DOTTORATO DI RICERCA
in
SCIENZE COMPUTAZIONALI E INFORMATICHE
Ciclo XXI

Consorzio tra Università di Catania, Università di Napoli Federico II,
Seconda Università di Napoli, Università di Palermo, Università di Salerno

SEDE AMMINISTRATIVA: UNIVERSITÀ DI NAPOLI FEDERICO II

MICHELANGELO BUCCI

GENERALIZATIONS OF EPISTURMIAN WORDS AND MORPHISMS

TESI DI DOTTORATO DI RICERCA

“*abaababa...*”

My nephew.

Ringraziamenti

Sin da piccolo ho sempre amato giocare con le parole. Quando scelsi di fare matematica, non avrei mai immaginato di poter continuare a coltivare questa mia passione anche all'università. I primi, dovuti e sentiti, ringraziamenti vanno dunque al prof. de Luca, che mi ha dato la possibilità di continuare a giocare, e ad Alessandro, mio amico, collega e compagno di giochi.

Ringrazio i miei genitori, i miei fratelli, mia cognata e tutta la mia famiglia, che mi hanno accompagnato in questo tri-quadriennale cammino di Dottorato.

Grazie (in rigoroso ordine alfabetico) a Ilaria, Valeria, Viola e tutte le mie insostituibili amiche, vicine e lontane, perché se non ci fossero sarei costretto a frequentare solo uomini e, credetemi, non è cosa da poco.

Ringrazio Rosario, Marco, i cantori della SPP in tutte le sue varie declinazioni e tutti coloro che hanno cantato e cantano con me, perché senza di loro rischierei persino di avere del tempo libero.

Infine, come sempre, ringrazio Silvia per gli ultimi nove anni e per tutti quelli che sceglierà di passare con me.

Contents

Introduction

Despite being in some sense a pure "mathematical" topic, Combinatorics on Words, that is the study of structural and combinatorial properties of sequences of symbols, has a variety of applications in many distinct fields which range from Computer Science, to Biology, Physics and many others (see for instance [44, 45, 46]), which further increases the intrinsic interest of the subject. In this context, the theory of the so-called Sturmian words and the study of their fascinating properties has always been of prime importance, with aspects related to various fields such as Number Theory, Geometry, Computer Vision, Symbolic Dynamics, Theoretical Physics, Crystallography and so on. The reason for such a wide variety of applications probably relies in the impressive number of possible equivalent definitions and characterizations for Sturmian words, an aspect which has been shown since the fundamental work of Morse and Hedlund (cf. [51]) in 1940. For instance, an infinite sequence of a's and b's is Sturmian if it is not periodic and balanced (which means that taken any two substrings of the same length, the number of a's in the two substrings can differ at most by 1). Sturmian words can also be characterized as the words having, for each length n, exactly $n+1$ factors of length n. In this sense, they are the "simplest" (with lowest *factor complexity*) infinite words which are not periodic; many interesting properties of Sturmian words can be attributed to their "simplicity". What is perhaps the most evident limit of Sturmian words is that they are words defined over a two letters alphabet. It then appears natural to extend, where possible, the definitions to any finite alphabet; in many cases such extension is straightforward, but, as one can expect, many of the modified definitions are no longer equivalent. Among the other extensions of Sturmian words, the class of Episturmian words play a central role. Episturmian words have in common with Sturmian words many interesting properties and are themselves a well studied subject of interest. Among the other possible definitions, since Episturmian words share with Sturmian words the property of having a great number of equivalent definitions, a word is said to be Episturmian if it is closed under reversal (meaning that if a w is a factor of our word, also the reversal of w must appear as a factor) and if it has at most one left special factor for each length (a left special factor is a factor which occurs preceded by at least two distinct letters). Once again, it appears quite natural to consider and study nontrivial extensions of Episturmian words, like it has been done for Sturmian words.

In this thesis two kind of extensions are considered: extension obtained weakening some hypothesis in the definitions of Episturmian words, and an extension obtained considering the class of all word having a property which Episturmian words enjoy, but by which they are not characterized. More specifically, starting from biological considerations, we considered the extension obtained substituting in two definitions the reversal operator with a generic involutory antimorphism (just to stress the meaning of the the operation, we recall that the Watson-Crick complementarity law is basically an involutory antimorphism); in this way we obtained two interesting and different extensions of the first kind. For what concerns the second kind of extensions, we studied two characterizing properties of words which have (like Episturmian words) a maximal number of palindromic factors.

All the extensions considered have many potential applications which are yet to be studied, and, since the topic is both promising and new, certainly deserve to be the subject of further analysis.

Overview

The first part of this thesis is dedicated to the study of the notions needed to introduce the generalizations and more advanced notions presented in the second part.

In Chapter 1, after some algebraic notions, the basic concepts and terminology of combinatorics on words are given.

Chapter 2 is devoted to the introduction of Sturmian and Episturmian words and morphisms, which are the base for the generalizations presented in the following chapters. The first section covers Sturmian words; though this section probably covers much more than what is effectively needed in the remaining of the thesis, the importance of the topic is such that we felt compelled to present at least an essential survey, with a look at some recent results like those on Sturmian palindromes. The following sections are dedicated to Episturmian words and morphisms and contain most of the definition that are extended in the following.

With Chapter 3, where the first generalizations of Episturmian words are presented, we get into the core of the thesis. In the first section of the chapter we introduce the idea on which the following work is based: that of substituting a generic involutory antimorphism to the reversal operator in the definitions used for Episturmian words. Following this idea, we introduce the derived notions of pseudopalindromes and the important class of unbordered pseudopalindromes. At this point we use these concepts to extend, in the three following sections, the equivalent definitions for Episturmian words given in Chapter 2. In section 3.5 we show what are the relations between the new classes of words introduced, showing that the definitions which are equivalent for the Episturmian case are no longer so and that ϑ-standard words and standard ϑ-Episturmian words are both proper subclasses of the ϑ-standard words with "seed". In the following section, we go back to the class of ϑ-Episturmian words (which are somewhat the most difficult class to deal with, since its definition is far from being constructive), showing several useful structural results and characterizing the standard ϑ-Episturmian words which are also ϑ-standard. Many of the results of this section will be useful in the following chapter. In Section 3.7, mainly based on [14], we show that the choice of considering involutory antimorphisms to generalize Episturmian words is actually a good choice, since under some weak and natural conditions, words turn out to be ϑ-Episturmian for some ϑ. That is, if we consider only some conditions which are always true for Episturmian words (plus some weak ones which are true for a wide class of Episturmian words), there must be some ϑ such that the words considered are ϑ-Episturmian. In the last section of the chapter, we show that a wider class of standard ϑ-Episturmian (i.e. dropping one of the requirements in the definition), still retains (like the "normal" ϑ-Episturmian words) a strong link with Arnoux-Rauzy words.

Chapter 4 is completely devoted to the study of characteristic morphisms, i.e., morphisms that map standard Episturmian words into standard ϑ-Episturmian words, which have been thoroughly studied in [11]. Such morphisms, which are interesting by themselves, are a useful mean of constructing ϑ-Episturmian words, which is not a straightforward task from the definition. The main result of this chapter is the complete and constructive characterization of characteristic morphisms given in Section 4.3, with its vast proof.

In Chapter 5, finally, a completely different generalization of Episturmian words is given; in this chapter we introduce the class of rich words, i.e., words which have a maximal number of palindromic factors, and give two characterizing properties of such words: the first is an equation which links the factor complexity (the function which counts, for each length, the number of factors of such length) of a rich word to its palindromic factor complexity; the second is a structural one and characterizes as palindromes all the factors beginning with a word, ending with its reversal and having no other occurences of either one in between.

Chapter 1

Preliminary notions and definitions

1.1 Basic algebraic notions

As is well known (see for instance [21]), a *semigroup* S is a set in which an associative binary operation (product) is defined.

A *monoid* M is a semigroup having an *identity* element 1_M such that $1_M x = x 1_M = x$ for all $x \in M$. A subsemigroup N of M is a *submonoid* if $1_M \in N$.

The product operation on a semigroup S can be naturally extended to the powerset $\mathcal{P}(S)$: given $X, Y \subseteq S$, we define

$$XY = \{xy \in S \mid x \in X \text{ and } y \in Y\} .$$

It is also common to define left and right *quotients*, by setting

$$X^{-1}Y = \{w \in S \mid Xw \cap Y \neq \emptyset\} ,$$

and

$$YX^{-1} = \{w \in S \mid wX \cap Y \neq \emptyset\} .$$

We shall often confuse singletons and their elements, when this does not lead to ambiguity. For instance, if $x \in S$ and $Y \subseteq S$, by xY we will mean the set $\{x\}Y$.

The *subsemigroup generated by* $X \subseteq S$ is the smallest subsemigroup of S containing X, and coincides with

$$X^+ = \bigcup_{n>0} X^n \ .$$

Similarly, the *submonoid generated by* $X \subseteq M$ is equal to

$$X^* = \bigcup_{n\geq 0} X^n \ ,$$

where conventionally $X^0 = \{1_M\}$.

Given two semigroups S, S', a *morphism* (resp. *antimorphism*) φ from S to S' is a map

$$\varphi : S \longrightarrow S'$$

such that $\varphi(xy) = \varphi(x)\varphi(y)$ (resp. $\varphi(xy) = \varphi(y)\varphi(x)$) for all $x, y \in S$. A monoid (anti-)morphism $\varphi : M \to M'$ is a semigroup (anti-)morphism such that $\varphi(1_M) = 1_{M'}$. An *isomorphism* is a bijective morphism, and an *automorphism* of M is an isomorphism between M and itself. When $\varphi : M \to M'$ is a morphism or antimorphism and $x \in M$, we shall often use the exponential notation x^φ for $\varphi(x)$.

A semigroup S (resp. monoid M) is *free over* $X \subseteq S$ (resp. $X \subseteq M$) if every element of X^+ admits a unique factorization over X, and $X^+ = S$ (resp. $X^* = M$). Free semigroups (monoids) over sets of the same cardinality are isomorphic.

1.2 Basic combinatorics on words

The free monoid of words

Let A be a nonempty finite set, or *alphabet*, whose elements are called *letters*. The set of finite sequences of letters, or *words* over A, can be naturally endowed with the binary operation of *concatenation*. The semigroup A^+ thus obtained is free over A: a word $w \in A^+$ can be written uniquely as a product of letters $w = a_1 a_2 \cdots a_n$, with $a_i \in A$, $i = 1, \ldots, n$. Therefore A^+ is called *the* free semigroup over A. The free monoid A^* is obtained by adding an identity element, the *empty word* $\varepsilon = 1_{A^*}$, to A^+: $A^* = A^+ \cup \{\varepsilon\}$.

Let $w = a_1 \cdots a_n \in A^+$, where $a_i \in A$ for $1 \leq i \leq n$. The integer n is the *length* of w, denoted by $|w|$. It is natural to set $|\varepsilon| = 0$.

A word u is a *factor* of $w \in A^*$ if $w = rus$ for some words r and s. In the special case $r = \varepsilon$ (resp. $s = \varepsilon$), u is called a *prefix* (resp. *suffix*) of w. A factor u of w is *proper* if $u \neq w$; it is *median* if $w = rus$ with $|r| = |s|$. We denote respectively by $\operatorname{Fact} w$, $\operatorname{Pref} w$, and $\operatorname{Suff} w$ the sets of all factors, prefixes, and suffixes of the word w.

A subset of A^* is often called a *language* over A. For $Y \subseteq A^*$, $\operatorname{Pref} Y$, $\operatorname{Suff} Y$, and $\operatorname{Fact} Y$ will denote respectively the languages of prefixes, suffixes, and factors of all the words of Y; in symbols,

$$\operatorname{Fact} Y = \bigcup_{w \in Y} \operatorname{Fact} w \ ,$$

and similarly for $\operatorname{Pref} Y$ and $\operatorname{Suff} Y$.

A *code* over A is a language $Z \subseteq A^*$ such that the monoid Z^* is free over Z. Thus Z is a code if and only if whenever $z_1, z_2, \ldots, z_n, z'_1, \ldots, z'_m \in Z$ are such that

$$z_1 \cdots z_n = z'_1 \cdots z'_m \ ,$$

then $n = m$ and $z_i = z'_i$ for $i = 1, \ldots, n$. A *prefix* (resp. *suffix*) *code* is a subset of A^+ with the property that none of its elements is a proper prefix (resp. suffix) of any other. Any prefix (or suffix) code is in fact a code. A *biprefix* code is a code which is both prefix and suffix.

1.2.1 Borders and periods

A factor of $w \in A^*$ is called a *border* of w if it is both a prefix and a suffix of w. A word is called *unbordered* if its only proper border is ε. Since the set of proper borders of the empty word is empty, coherently with the above definition we do not consider ε unbordered.

A positive integer p is a *period* of $w = a_1 \cdots a_n$ ($a_i \in A$, $i = 1, \ldots, n$) if whenever $1 \leq i, j \leq |w|$ one has that

$$i \equiv j \pmod{p} \Longrightarrow a_i = a_j \ .$$

Note that with this definition, any $n \geq |w|$ is a period of w. As is well known and quite evident (cf. [44]), a word w has a period $p \leq |w|$ if and only if it has a border of length $|w| - p$. We denote by π_w the minimal period of w, and set $\pi_\varepsilon = 1$. Thus a word w is unbordered if and only if $\pi_w = |w|$. If w is nonempty, then its *fractional root* z_w is its prefix of length $|z_w| = \pi_w$. We can write any nonempty word w as

$$w = z_w^k z'$$

where z_w is the fractional root of w, the integer $k \geq 1$ is sometimes called the *order* of w, and z' is a proper prefix of z_w.

We recall the following fundamental result about periodicity (cf. [44]):

Theorem 1.2.1 (Fine and Wilf). *If a word w has two periods p and q, and $|w| \geq p + q - \gcd(p, q)$, then w has also the period $\gcd(p, q)$.*

1.2.2 Infinite words and limits

An *infinite word* (from left to right) x over the alphabet A is just an infinite sequence of letters, i.e., a mapping $x : \mathbb{N}_+ \longrightarrow A$ where $\mathbb{N}_+$ is the set of positive integers. One can represent x as

$$x = x_1 x_2 \cdots x_n \cdots ,$$

where for any $i > 0$, $x_i = x(i) \in A$. A (finite) *factor* of x is either the empty word or any sequence $x_i \cdots x_j$ with $i \leq j$, i.e., any block of consecutive letters of x. If $i = 1$, then u is a *prefix* of x. We denote by $\operatorname{Fact} x$ and $\operatorname{Pref} x$ the sets of finite factors and prefixes of x respectively.

The product between a finite word w and an infinite one x is naturally defined as the infinite word wx having w as a prefix and $x_{j-|w|}$ as its j-th letter, for all $j > |w|$. The set of all infinite words over A is denoted by A^ω. We also set $A^\infty = A^* \cup A^\omega$.

A metric on A^ω can be defined by setting $d(x, x) = 0$ for $x \in A^\omega$, and

$$d(x, y) = 2^{-\ell}$$

for $y \neq x$, where $\ell = \max\{n \in \mathbb{N} \mid \operatorname{Pref} x \cap \operatorname{Pref} y \cap A^n \neq \emptyset\}$ is the length of the maximal common prefix of x and y. This metric induces the product topology on $A^\omega = A^{\mathbb{N}_+}$ (where A is discrete), making it a compact, perfect, and totally disconnected metric space, that is, a *Cantor space* (cf. [50]). The metric d can be "extended" to the whole A^∞ in the following way: define (as above) the metric d' on $(A')^\omega$, where $A' = A \cup \{\$\}$ and $\$ \notin A$; then identify any $w \in A^*$ with the infinite word $w\$^\omega$. In this way A^∞ is regarded as a subspace of $(A')^\omega$.

The main benefit of topology for our purposes is the possibility of taking limits of sequences. We recall that convergence with respect to the product topology is *pointwise*, so that a sequence of words $(z_m)_{m\geq 0}$ in A^∞ converges to an infinite word $x = x_1 \cdots x_n \cdots$ if and only if for any $k > 0$, there exists some $N \geq 0$ such that for all $n \geq N$, the k-th letter of z_n exists (i.e., $z_n \in A^\omega$ or $|z_n| \geq k$) and is equal to x_k. For instance, the sequence

$$(a^m b)_{m\geq 0}$$

converges to the infinite word $a^\omega = aaa\cdots$. A wide family of convergent sequences, which will appear frequently in the following chapters, is made of all sequences of finite words $(z_m)_{m\geq 0}$ such that for sufficiently large n, the word z_n is a prefix of z_{n+1}.

For any $Y \subseteq A^*$, Y^ω denotes the set of infinite words which can be factorized by the elements of Y. The above example shows that an infinite word which is the limit of a sequence of words of Y^* need not be in Y^ω (take $Y = a^*b$); however, it is in Y^ω if Y is finite.

1.2.3 Further definitions and properties

Let $w \in A^\infty$. An *occurrence* of a factor u in w is any pair $(\lambda, \rho) \in A^* \times A^\infty$ such that $w = \lambda u \rho$. If $a \in A$ and $w \in A^*$, $|w|_a$ denotes the number of occurrences of a in the word w; trivially we have

$$|w| = \sum_{a\in A} |w|_a \, .$$

For $w \in A^\infty$, $\operatorname{alph} w$ denotes the set of letters occurring in w, that is, $\operatorname{alph} w = \{a \in A \mid |w|_a > 0\}$.

Let $s \in A^\infty$ and $w, u \in \operatorname{Fact} s$. We call w a *first return* to u in s if w contains exactly two distinct occurrences of u, one as a prefix and the other as a suffix, i.e.,

$$w = u\lambda = \mu u \quad \text{with } \lambda, \mu \in A^+ \text{ and } w \notin A^+ u A^+ \, .$$

We observe that in such a case, $wu^{-1} = \mu$ is usually called a *return word over* u in s (see [31]). We call the integer $|\mu|$ the *shift* of the first return. An infinite word s is said *uniformly recurrent* if for any $v \in \operatorname{Fact} s$, the shifts of the first returns to v in s are bounded above by a constant c_v.

If $x \in A$ and vx (resp. xv) is a factor of $w \in A^\infty$, then vx (resp. xv) is called a *right* (resp. *left*) *extension* of v in w. We recall that a factor v of a (finite or infinite) word w is called *right special* if it has at least two distinct right extensions in w, i.e., there exist at least two distinct letters $a, b \in A$ such that both va and vb are factors of w. *Left special* factors are defined analogously. A factor of w is called *bispecial* if it is both right and left special.

We denote by R_w the smallest integer k, if it exists, such that w has no right special factor of length k (and we set $R_w = \infty$ otherwise, that is, when w is an infinite word having arbitrarily long right special factors). The following noteworthy inequality (cf. [25]) relates the minimal period π_w of a finite word w and R_w:

$$\pi_w \geq R_w + 1 \, . \tag{1.1}$$

Symmetrically, one can introduce the parameter L_w as the minimal length for which w has no *left* special factors; L_w satisfies $\pi_w \geq L_w + 1$ too.

A finite word w is *primitive* if it cannot be written as a power $w = u^k$ with $k > 1$. Clearly any unbordered word is primitive, but the converse is false: consider for instance the word aba. We denote by $\pi(A^*)$ the set of all primitive words over A. As is well known (cf. [44]), for any nonempty word w there exists a unique primitive word u such that $w = u^k$ for some $k \geq 1$. Such a u is usually called the *(primitive) root* of w and denoted by $\sqrt{w}$.

Two words $u, v \in A^*$ are *conjugate* if there exist $\lambda, \mu \in A^*$ such that $u = \lambda\mu$ and $v = \mu\lambda$. Conjugacy is an equivalence relation in A^*; we write $u \sim v$ if u and v are conjugate.

Suppose that $\leq$ is a total order on A. One can extend this order to the *lexicographic* order on A^* by letting, for all $v, w \in A^*$,

$$v \leq w \iff (v \in \operatorname{Pref} w \ \text{ or } \ v = uav',\, w = ubw') \, ,$$

for some $u, v', w' \in A^*$ and $a, b \in A$ such that $a < b$.

A word is called a *Lyndon* (resp., *anti-Lyndon*) word if it is primitive and minimal (resp., maximal) in its conjugacy class, with respect to the lexicographic order. For instance, if $a < b$ then $w = aabab$ is a Lyndon word, for its conjugates ($ababa$, $babaa$, $abaab$, and $baaba$) are all lexicographically greater than w.

In the sequel, we shall need the two following simple lemmas; we report the proofs for the sake of completeness.

Lemma 1.2.2. *A word $w \in A^*$ has the period $p \leq |w|$ if and only if all its factors having length p are in the same conjugacy class.*

Proof. The case $w = \varepsilon$ is trivial. Then suppose that p is a period of $w = a_1 \cdots a_n$, $a_i \in A$, $i = 1, \ldots, n$. Let u be a factor of w of length p. By the definition of period, there exists a positive integer $i \leq p$ such that $u = a_i a_{i+1} \cdots a_p a_1 a_2 \cdots a_{i-1}$, so that u is a conjugate of $a_1 a_2 \cdots a_p$.

The converse is an easy consequence of the following fact: if $x, y \in A$ and $u \in A^*$, then $xu \sim uy$ if and only if $x = y$. Therefore, if all factors of w of length p are conjugate, one derives that $a_i = a_{i+p}$ for all i such that $1 \leq i \leq n - p$. □

Lemma 1.2.3. *A word $w \in A^*$ is primitive if and only if $\pi_{w^k} = |w|$ for any integer $k \geq 2$.*

Proof. Let w be a primitive word, and suppose that w^k has a period $q \leq |w|$. Since $|w|$ is a period of w^k and $|w^k| = k|w| > |w| + q$, by Theorem 1.2.1, w^k, as well as w, has also the period $d = \gcd(q, |w|)$. Thus $w = u^{|w|/d}$ for some u; this implies $|w|/d = 1$ and then $q = |w|$, as w is primitive.

Conversely, suppose $w \in A^*$ is not primitive. If $w = \varepsilon$, then

$$\pi_{w^k} = \pi_\varepsilon = 1 \neq 0 = |w| \ .$$

Let then $w \in A^+$ and let u be its primitive root. Clearly $|u|$ is a period of w^k, and $|u| < |w|$. □

We remark that also the *fractional* root z_w of a nonempty word w is trivially primitive. Hence, by Lemma 1.2.3 we obtain that for any $w \in A^+$ and $k \geq 2$,

$$\pi_w = \pi_{z_w^k} \ . \tag{1.2}$$

We remark that, by symmetrical arguments, one can prove results analogous to Proposition 2.1.13 and Theorem 2.1.37, namely, *if $\pi_w = L_w + 1$, then $w \in St$*, and *a palindrome $w \in A^*$ is Sturmian if and only if $\pi_w = L_w + 1$.*

1.3 Overlap-free and normal codes

We say that a code Z over A is *overlap-free* if no two of its elements overlap properly, i.e., if for all $u, v \in Z$, $\text{Suff}\, u \cap \text{Pref}\, v \subseteq \{\varepsilon, u, v\}$.

For instance, let $Z_1 = \{a, bac, abc\}$ and $Z_2 = \{a, bac, cba\}$. One has that Z_1 is an overlap-free and suffix code, whereas Z_2 is a prefix code which is not overlap-free as bac and cba overlap properly.

A code $Z \subseteq A^+$ will be called *right normal* if it satisfies the following condition:

$$(\text{Pref}\, Z \setminus Z) \cap RS\, Z \subseteq \{\varepsilon\}\ , \tag{1.3}$$

i.e., any proper and nonempty prefix u of any word of Z such that $u \notin Z$ is not right special in Z. In a symmetric way, a code Z is called *left normal* if it satisfies the condition

$$(\text{Suff}\, Z \setminus Z) \cap LS\, Z \subseteq \{\varepsilon\}\ . \tag{1.4}$$

A code Z is called *normal* if it is right and left normal.

As an example, the code $Z_1 = \{a, ab, bb\}$ is right normal but not left normal; the code $Z_2 = \{a, aba, aab\}$ is normal. The code $Z_3 = \{a, cad, bacadad\}$ is biprefix, overlap-free, and right normal, and the code $Z_4 = \{a, badc\}$ is biprefix, overlap-free, and normal.

In the rest of this section, we analyse some properties of left (or right) normal codes, under some additional requirements such as being suffix, prefix, or overlap-free. We stress that all statements of the following propositions can be applied to codes which are biprefix, overlap-free, and normal.

A first noteworthy result, which will be useful in the sequel, is the following:

Proposition 1.3.1. *Let Z be a biprefix, overlap-free, and right normal (resp. left normal) code. Then:*

1. *if $z \in Z$ is such that $z = \lambda v \rho$, with $\lambda, \rho \in A^*$ and v a nonempty prefix (resp. suffix) of $z' \in Z$, then $\lambda z'$ (resp. $z' \rho$) is a prefix (resp. suffix) of z, proper if $z \neq z'$.*

2. *for $z_1, z_2 \in Z$, if $z_1^f = z_2^f$ (resp. $z_1^\ell = z_2^\ell$), then $z_1 = z_2$.*

Proof. Let $z = \lambda v \rho$ with $v \in \mathrm{Pref}\, z'$ and $v \neq \varepsilon$. If $v = z'$, there is nothing to prove. Suppose then that v is a proper prefix of z'. Since Z is a prefix code, any proper nonempty prefix of z', such as v, is not an element of Z; moreover, it is not right special in Z, since Z is right normal. Therefore, to prove the first statement it is sufficient to show that $|v\rho| \geq |z'|$, where the inequality is strict if $z \neq z'$. Indeed, if $|v\rho| < |z'|$, then a proper prefix of z' would be a suffix of z, which is impossible as Z is an overlap-free code. If $|v\rho| = |z'|$, then $z' \in \mathrm{Suff}\, z$, so that $z' = z$ as Z is a suffix code.

Let us now prove the second statement. Let $z_1, z_2 \in Z$ with $z_1^f = z_2^f$. By contradiction, suppose $z_1 \neq z_2$. By the preceding statement, we derive that z_1 is a proper prefix of z_2 and z_2 is a proper prefix of z_1, which is clearly absurd. The symmetrical claims can be analogously proved. □

From the preceding proposition, a biprefix, overlap-free, and normal code satisfies both properties 1 and 2 and their symmetrical statements.

The following general lemma on prefix codes, will be very useful in the next sections:

Lemma 1.3.2. *Let $g : B^* \to A^*$ be an injective morphism such that $g(B) = Z$ is a prefix code. Then for all $p \in B^*$ and $q \in B^\infty$ one has that p is a prefix of q if and only if $g(p)$ is a prefix of $g(q)$.*

Proof. The 'only if' part is trivial. Therefore, let us prove the 'if' part. Let us first suppose $q \in B^*$, so that $g(q) = g(p)\zeta$ for some $\zeta \in A^*$. Since $g(p), g(q) \in Z^*$ and Z^* is left unitary, it follows that $\zeta \in Z^*$. Therefore, there exists, and is unique, $r \in B^*$ such that $g(r) = \zeta$. Hence $g(q) = g(p)g(r) = g(pr)$. Since g is injective one has $q = pr$ which proves the assertion in this case. If $q \in B^\omega$, there exists a prefix $q_{[n]}$ of q such that $g(p) \in \mathrm{Pref}\, g(q_{[n]})$. By the previous argument, it follows that p is a prefix of $q_{[n]}$ and then of q. □

Lemma 1.3.3. *Let Z be a left normal and suffix code over A. For any $a, b \in A$, $a \neq b$, $\lambda \in A^+$, if $a\lambda, b\lambda \in \mathrm{Fact}\, Z^*$ and $\lambda \notin \mathrm{Pref}\, Z^*$, then $a\lambda, b\lambda \in \mathrm{Fact}\, Z$.*

Proof. By symmetry, it suffices to prove that $a\lambda \in \mathrm{Fact}\, Z$. By hypothesis there exist words $v, \zeta \in A^*$ such that $va\lambda\zeta = z_1 \cdots z_n$, with $n \geq 1$ and $z_i \in Z$,

$i = 1, \ldots, n$. If $n = 1$, then $a\lambda \in \operatorname{Fact} Z$ and we are done. Then suppose $n > 1$, and write:

$$va = z_1 \cdots z_h \delta, \quad \delta\lambda\zeta = z_{h+1} \cdots z_n, \quad z_{h+1} = \delta\xi = z\ , \tag{1.5}$$

with $\delta \in A^*$, $h \geq 0$, and $\xi \neq \varepsilon$. Let us observe that $\delta \neq \varepsilon$, for otherwise $\lambda \in \operatorname{Pref} Z^*$, contradicting the hypothesis on λ.

If $|\delta\lambda| \leq |z|$, then since $a = \delta^\ell$, we have $a\lambda \in \operatorname{Fact} Z$ and we are done. Therefore, suppose $|\delta\lambda| > |z|$. This implies that ξ is a proper prefix of λ, and by (1.5), a proper suffix of z. Moreover, as $a = \delta^\ell$, we have $a\xi \in \operatorname{Fact} Z$.

Since $b\lambda \in \operatorname{Fact} Z^*$, in a symmetric way one derives that either $b\lambda \in \operatorname{Fact} Z$, or there exists $\xi' \neq \varepsilon$ which is a proper prefix of λ and a proper suffix of a word $z' \in Z$. In the first case we have $b\lambda \in \operatorname{Fact} Z$, so that $a\xi, b\xi \in \operatorname{Fact} Z$, whence $\xi \in \operatorname{Suff} Z \cap LS\, Z$, and $\xi \notin Z$ since Z is a suffix code. We reach a contradiction since $\xi \neq \varepsilon$ and Z is left normal.

In the second case, ξ and ξ' are both prefixes of λ. Let $\hat{\xi}$ be in $\{\xi, \xi'\}$ with minimal length. Then $a\hat{\xi}, b\hat{\xi} \in \operatorname{Fact} Z$, so that $\hat{\xi} \in \operatorname{Suff} Z \cap LS\, Z$. Since $\hat{\xi} \notin Z$, as Z is a suffix code, we reach again a contradiction because $\hat{\xi} \neq \varepsilon$ and Z is left normal. Therefore, the only possibility is that $a\lambda \in \operatorname{Fact} Z$. □

Proposition 1.3.4. *Let Z be a suffix, left normal, and overlap-free code over A, and let $a, b \in A$, $v \in A^*$, $\lambda \in A^+$ be such that $a \neq b$, $va \notin Z^*$, $va\lambda \in \operatorname{Pref} Z^*$, and $b\lambda \in \operatorname{Fact} Z^*$. Then $a\lambda \in \operatorname{Fact} Z$.*

Proof. Since $va\lambda \in \operatorname{Pref} Z^*$, there exists $\zeta \in A^*$ such that $va\lambda\zeta = z_1 \cdots z_n$, $n \geq 1$, $z_i \in Z$, $i = 1, \ldots, n$. Then we can assume that (1.5) holds for suitable $h \geq 0$, $\delta \in A^*$, and $\xi \in A^+$. We have $n > 1$, for otherwise the statement is trivial, and $\delta \neq \varepsilon$ since $va \notin Z^*$. As $\delta^\ell = a$, if $|\delta\lambda| \leq |z|$ we obtain $a\lambda \in \operatorname{Fact} Z$ and we are done. Therefore assume $|\delta\lambda| > |z|$. In this case ξ is a proper prefix of λ and a proper suffix of z. If $\lambda \in \operatorname{Pref} Z^*$ we reach a contradiction, since $\xi \in \operatorname{Suff} Z \cap \operatorname{Pref} Z^*$ and this contradicts the hypothesis that Z is a suffix and overlap-free code. Thus $\lambda \notin \operatorname{Pref} Z^*$; this implies, by the previous lemma, that $a\lambda \in \operatorname{Fact} Z$. □

Proposition 1.3.5. *Let Z be a biprefix, overlap-free, and right normal code over A. If $\lambda \in \operatorname{Pref} Z^* \setminus \{\varepsilon\}$, then there exists a unique word $u = z_1 \cdots z_k$*

with $k \geq 1$ *and* $z_i \in Z$, $i = 1, \ldots, k$, *such that*

$$u = z_1 \cdots z_k = \lambda\zeta, \quad z_1 \cdots z_{k-1}\delta = \lambda\ , \tag{1.6}$$

where $\delta \in A^+$ *and* $\zeta \in A^*$.

Proof. Let us suppose that there exist $h \geq 1$ and words $z'_1, \ldots, z'_h \in Z$ such that

$$z'_1 \cdots z'_h = \lambda\zeta', \quad z'_1 \cdots z'_{h-1}\delta' = \lambda \tag{1.7}$$

with $\zeta' \in A^*$ and $\delta' \in A^+$. From (1.6) and (1.7) one obtains $u = z_1 \cdots z_k = z'_1 \cdots z'_{h-1}\delta'\zeta$ and $z'_1 \cdots z'_h = z_1 \cdots z_{k-1}\delta\zeta'$, with $z_k = \delta\zeta$ and $z'_h = \delta'\zeta'$. Since Z is a biprefix code, we derive $h = k$ and consequently $z_i = z'_i$ for $i = 1, \ldots, k-1$. Indeed, if $h \neq k$, we would derive by cancellation that $\delta'\zeta = \varepsilon$ or $\delta\zeta' = \varepsilon$, which is absurd as $\delta, \delta' \in A^+$.

Hence we obtain $z_k = \delta'\zeta = \delta\zeta$, whence $\delta = \delta'$. Thus δ is a common nonempty prefix of z_k and z'_k. Since Z is right normal, by Proposition 1.3.1 we obtain that z_k is a prefix of z'_k and *vice versa*, i.e., $z_k = z'_k$. □

Proposition 1.3.6. *Let* Z *be a biprefix, overlap-free, and normal code over* A. *If* $u \in Z^* \setminus \{\varepsilon\}$ *is a proper factor of* $z \in Z$, *then there exist* $p, q \in Z^*$, $h, h' \in A^+$ *such that* $h^\ell \notin \operatorname{Suff} Z$, $(h')^f \notin \operatorname{Pref} Z$, *and*

$$z = hpuqh'\ .$$

Proof. Since u is a proper factor of $z \in Z$, there exist $\xi, \xi' \in A^*$ such that $z = \xi u \xi'$; moreover, ξ and ξ' are both nonempty as Z is a biprefix code. Let p (resp. q) be the longest word in $\operatorname{Suff} \xi \cap Z^*$ (resp. $\operatorname{Pref} \xi' \cap Z^*$), and write

$$z = \xi u \xi' = hpuqh'$$

for some $h, h' \in A^*$. Since u and hp are nonempty and Z is a biprefix code, one derives that h and h' cannot be empty. Moreover, $h^\ell \notin \operatorname{Suff} Z$ and $(h')^f \notin \operatorname{Pref} Z$, for otherwise the maximality of p and q would be contradicted using Proposition 1.3.1. □

Chapter 2

Sturmian and Episturmian words

2.1 Sturmian words

Sturmian words were first considered in the 18th century by J. Bernoulli III, in his astronomical studies. Several authors later developed the subject from different points of view, but the first systematic study was given in 1940 by M. Morse and G. A. Hedlund (cf. [51]). They were also the first to use the name Sturmian, in honor of C. F. Sturm.

By definition, an infinite word is *Sturmian* if for each $n \in \mathbb{N}$ it has $n+1$ distinct factors of length n. This implies that a Sturmian word is on a two-letter alphabet, that will be $\mathcal{A} = \{a, b\}$ for the rest of this chapter (we shall keep using a non-calligraphic A for a generic alphabet). As is well known [45], an infinite binary word x is Sturmian if and only if for any $n \geq 0$ there is only one right special factor of x of length n.

A famous theorem by Morse and Hedlund (cf. [50]) states that an infinite word s has less than $n+1$ factors for some $n \geq 0$ if and only if it is *eventually periodic*, that is, writable as $s = uv^{\omega}$ for some finite words u, v. Thus Sturmian words have the smallest possible number of factors of each length, among all infinite words which are not eventually periodic.

A first description of the structure of Sturmian words was given in [51], where the following well-known characterization is found: an infinite word $s \in \mathcal{A}^{\omega}$ is Sturmian if and only if it is not eventually periodic and it is *balanced*, i.e., it satisfies, for all $n \geq 0$ and $u, v \in \mathcal{A}^n \cap \mathrm{Fact}\, s$,

$$||u|_a - |v|_a| \leq 1 . \tag{2.1}$$

2.1.1 Standard and central Sturmian words

An equivalent geometrical definition of Sturmian words can be given in terms of *cutting sequences*. In fact, a Sturmian word can be defined by considering the sequence of cuts in a squared lattice ($\mathbb{N} \times \mathbb{N}$) made by a ray having a slope which is an irrational number α. A horizontal cut is denoted by the letter b, a vertical by a, and a cut with a corner by ab or ba.

A Sturmian word represented by a ray starting from the origin is usually called *standard* or *characteristic*. We shall denote by c_α the standard Sturmian word associated with the irrational slope α. Standard Sturmian words can be equivalently defined as follows. For any sequence $d_0, d_1, \ldots, d_n, \ldots$ of integers such that $d_0 \geq 0$ and $d_i > 0$ for $i > 0$, one defines, inductively, the sequence of words $(s_n)_{n \geq 0}$ where

$$s_0 = b,\ s_1 = a,\ \text{ and } s_{n+1} = s_n^{d_{n-1}} s_{n-1},\ \text{ for } n \geq 1 . \tag{2.2}$$

The sequence $(s_n)_{n\geq 0}$ converges to a limit s which is an infinite standard Sturmian word. More precisely, one has $s = c_\alpha$, where the slope α is given by the continued fraction

$$\alpha = \cfrac{1}{d_0 + \cfrac{1}{d_1 + \cfrac{1}{\ddots}}} = [0; d_0, d_1, \ldots]$$

(see for instance [45]). Any standard Sturmian word can be generated in this way. If $d_i = 1$ for all $i \geq 0$, one obtains the famous *Fibonacci word*

$$f = abaababaabaababaababaa \cdots ,$$

whose slope is the inverse of the golden ratio.

We shall denote by *Stand* the set of all the words s_n, $n \geq 0$ of any sequence $(s_n)_{n\geq 0}$ constructed by the previous rule (2.2). Any word of *Stand* is called *finite standard (Sturmian) word*. We recall the following characterization of $Stand$ given in [28]:

$$Stand = \mathcal{A} \cup (PAL^2 \cap PAL\{ab, ba\}) \ , \tag{2.3}$$

i.e., a word $w \in \mathcal{A}^*$ is standard if and only if it is a letter or it satisfies the following equation:

$$w = \alpha\beta = \gamma xy \ ,$$

with $\alpha, \beta, \gamma \in PAL$ and $\{x, y\} = \mathcal{A}$.

A finite word w is called *central* if it has two periods p and q such that $\gcd(p, q) = 1$ and $|w| = p + q - 2$. Conventionally, the empty word ε is central (in this case, $p = q = 1$). Central words are over a two-letter alphabet. The set of all central words over $\mathcal{A} = \{a, b\}$ is usually denoted by PER. It is well known (see [28, 45]) that the set PER coincides with the set of palindromic prefixes of all standard Sturmian words. In the remaining part of this section we recall some properties of standard and central words which will be useful in the sequel.

The following important characterization of central words holds (see for instance [18]):

Proposition 2.1.1. *A word w is central over $\mathcal{A}$ if and only if w is a power of a letter of $\mathcal{A}$ or it satisfies the equation*

$$w = w_1abw_2 = w_2baw_1$$

for some words w_1 and w_2. Moreover, in this latter case, w_1 and w_2 are central words over $\mathcal{A}$, $p = |w_1| + 2$ and $q = |w_2| + 2$ are coprime periods of w, and $\min\{p, q\}$ is the minimal period of w.

Example 2.1.2. Let $w = aabaabaa \in PER$. We have

$$w = a(ab)aabaa = aabaa(ba)a \ ,$$

with $3 = \pi_w = |a| + 2$ and $7 = |aabaa| + 2$ being coprime periods of w, and $|w| = 8 = 3 + 7 - 2$.

From (2.3) and the preceding proposition, one easily derives (cf. [28]) that

$$Stand = \mathcal{A} \cup PER\{ab, ba\} \ , \tag{2.4}$$

i.e., any finite standard Sturmian word which is not a single letter is obtained by appending ab or ba to a central word. Conversely, any central word is obtained by deleting the last two letters of a standard word.

Let St be the set of finite Sturmian words, i.e., factors of infinite Sturmian words over the alphabet $\mathcal{A} = \{a, b\}$. We recall that for any infinite Sturmian word there exists an infinite standard Sturmian word having the same set of factors (cf. [45]). Therefore one easily derives that

$$St = \mathrm{Fact}(Stand) = \mathrm{Fact}(PER)\ . \tag{2.5}$$

Lemma 2.1.3 (see [18]). *If a central word w has the factor x^n, with $x \in \mathcal{A}$ and $n > 0$, then x^{n-1} is a prefix (and suffix) of w.*

Proposition 2.1.4 (see [52]). *A word w is central if and only if wab and wba are conjugate.*

Now let us suppose that the alphabet $\mathcal{A}$ is totally ordered by setting $a < b$.

Proposition 2.1.5 (see [6]). *The set $\mathcal{A} \cup aPERb$ is equal to the set of all Lyndon words which are Sturmian. Similarly, $\mathcal{A} \cup bPERa$ is the set of anti-Lyndon Sturmian words.*

Proposition 2.1.6 (see [39]). *A Sturmian word is unbordered if and only if it is a Lyndon or anti-Lyndon word.*

From Propositions 2.1.4 and 2.1.5, one derives the following interesting characterization of words conjugate of a standard word.

Proposition 2.1.7. *A primitive word $z \notin A$ is a conjugate of a standard word if and only if the Lyndon and the anti-Lyndon words in its conjugacy class have the same proper median factor of maximal length.*

Proof. Let z be a primitive word of length $|z| > 1$. Let s be a standard word conjugate to z. By (2.4), s can be written as $s = vxy$, with $v \in PER$ and $\{x, y\} = \mathcal{A}$. By Proposition 2.1.4, one derives that z is a conjugate of avb and bva. From Proposition 2.1.5, avb and bva are, respectively, a Lyndon and an anti-Lyndon word, so that the necessity is proved.

Conversely, let $z \in A^*$ and suppose that the Lyndon and the anti-Lyndon words in the conjugacy class of z can be written respectively as atb and bta, with $a, b \in A$ and $a < b$. By Proposition 2.1.4, one has that $t \in PER$, so that by (2.4), z is a conjugate of $tab \in Stand$. $\square$

2.1.2 Finite Sturmian words and periodicity

In this section we give two characterizations of finite Sturmian words, based on properties of their fractional root. We need some preliminary propositions. The first one gives some characterizations of the words w such that $w^2 \in St$ (such words have been called *cyclic balanced* in [20]). The equivalence of some of the conditions in Proposition 2.1.8 has recently been proved in [20] (see also [41]). We report here a more direct and simple proof for the sake of completeness.

Proposition 2.1.8. *Let w be a word. The following conditions are equivalent:*

1. $w^2 \in St$,
2. $w^* \subseteq St$,
3. *every conjugate of w^2 is Sturmian,*
4. *every conjugate of w is Sturmian,*
5. *the primitive root of w is a conjugate of a standard Sturmian word.*

Proof. *1.* $\Rightarrow$ *2.* Let $n > 2$. Any two factors of w^n of length $k > |w|/2$ overlap, thus it suffices to verify the balance condition only for factors of w^n of length $k \leq |w|/2$, which is satisfied because such words are also factors of $w^2 \in St$.

2. $\Rightarrow$ *3.* This is trivial, since any conjugate of w^2 is a factor of w^3.

3. $\Rightarrow$ *4.* This is trivial too, because the square of a conjugate of w is just a conjugate of w^2.

4. $\Rightarrow$ *5.* Let u be the primitive root of w. If every conjugate of w is Sturmian, then so is every conjugate of u. Hence it suffices to prove that if w is primitive, then it has a conjugate which is a standard word. Indeed, there exists a unique conjugate of w which is a Lyndon word, say w'. Since w' is Sturmian, by Proposition 2.1.5 one has that w' is either a letter or a word avb with $v \in PER$. In the former case, the desired standard conjugate is w' itself; in the latter case, one can take vba.

5. $\Rightarrow$ *1.* Let u be the primitive root of $w = u^k$; if v is a standard word in its conjugacy class, from equations (2.2) and (2.5) one derives that $v^2 \in St$.

Since *1.* $\Rightarrow$ *3.* and u^2 is a conjugate of v^2, one has $u^2 \in St$. As *1.* $\Rightarrow$ *2.*, this implies $w^2 = u^{2k} \in St$. $\square$

Let $w, u \in A^*$ with w unbordered; the word wu is called a *Duval extension* of w if no unbordered factor of wu is longer than w.

Proposition 2.1.9 (see [49]). *Every Duval extension wu of a Sturmian unbordered word w has the period $|w|$.*

We are now in the position of giving our first characterization of finite Sturmian words.

Theorem 2.1.10. *A nonempty word is Sturmian if and only if its fractional root is a conjugate of a standard word.*

Proof. Let w be a word. If its fractional root z_w is a conjugate of a standard word, then by Proposition 2.1.8, $z_w^* \subseteq St$, so that $w \in \operatorname{Fact} z_w^* \subseteq St$.

Conversely, let s be an unbordered factor of $w \in St$ of maximal length. One has $w = usv$ for suitable $u, v \in \mathcal{A}^*$. The word sv is a Duval extension of s, by the maximality of s. Since $\tilde{s}$ is unbordered too, and again by the maximality of s, the word $\tilde{s}\tilde{u} = \widetilde{us}$ is a Duval extension of $\tilde{s}$. From Proposition 2.1.9, one gets that both sv and $\widetilde{us}$ have the period $|s|$. This implies that also us has the period $|s|$.

By Lemma 1.2.2, all factors of us and sv having length $|s|$ are conjugates of s. Since any factor of w of length $|s|$ is either a factor of us or of sv, and s is a factor of both, we deduce from Lemma 1.2.2 that the whole w has the period $|s|$. Moreover, such period is minimal, because

$$|s| = \pi_s \leq \pi_w \leq |s| \ .$$

By Lemma 1.2.2, z_w is a conjugate of s; since s is an unbordered Sturmian word, by Proposition 2.1.6 it is a Lyndon (or anti-Lyndon) word, and therefore, by Proposition 2.1.5 it is in the set $\mathcal{A} \cup aPERb \cup bPERa$. Hence s, as well as z_w, is a conjugate of a standard word, which proves the assertion. $\square$

Examples 2.1.11. Let w be the word $aababaa$. Its fractional root $z_w = aabab$ is a conjugate of the standard word $ababa$, so that w is Sturmian.

Let $r = baabb$. In the conjugacy class of its root $z_r = baab$ there is no standard word, so that r is not Sturmian.

Corollary 2.1.12. *Let w be a nonempty word and z_w be its fractional root. Then w is a finite Sturmian word if and only if so is z_w^2.*

Proof. This is a straightforward consequence of the preceding theorem and of Proposition 2.1.8. ☐

The following proposition improves upon a result in [25].

Proposition 2.1.13. *Let w be a word. If $\pi_w = R_w + 1$, then w is Sturmian.*

Proof. Let $w \in A^*$. If $\pi_w = 1$, the result is trivially true. Thus we assume $\pi_w = R_w + 1 > 1$, so that there exists a right special factor s of w such that $|s| = \pi_w - 2$. Hence, there exist letters $a, b \in A$ such that $a \neq b$ and $sa, sb \in \operatorname{Fact} w$. The words sa and sb cannot be both suffixes of w, so we suppose, without loss of generality, that sa is not. Therefore one has either $saa \in \operatorname{Fact} w$ or $sac \in \operatorname{Fact} w$ with $c \neq a$. Since $|saa| = |sac| = \pi_w$, these two possibilities imply, respectively:

$$w \in \operatorname{Fact}((saa)^*) \tag{2.6}$$

or

$$w \in \operatorname{Fact}((sac)^*)\ . \tag{2.7}$$

We first show that (2.6) cannot hold. By contradiction, assume that it does hold. Since sb is a factor of w, it has to be a factor of $saas$ as well. We clearly have $sb \neq sa$, thus there exist $u, v \in A^*$ and $x \in A$ such that $saas = uxsbv$. The words u and v are respectively a prefix and a suffix of s, and $|u| + |v| = |saas| - |xsb| = 2|s| + 2 - |s| - 2 = |s|$. Therefore $s = uv$ and $vaau = xuvb$. But this is a contradiction, because $|vaau|_a > |xuvb|_a$.

Equation (2.7) is then satisfied. Let $u = sacsa$. The word $sb \in \operatorname{Fact} w$ has to be a factor of u; since sb is not a suffix of u, one has either $sba \in \operatorname{Fact} u$ or $sbx \in \operatorname{Fact} u$, with $x \neq a$. By Lemma 1.2.2, the latter is impossible, because $|sac| = |sbx| = \pi_w$ is a period of u, and $|sac|_a > |sbx|_a$. Thus sba is a factor of u, and by Lemma 1.2.2 it is a conjugate of sac. Therefore $c = b$; by Proposition 2.1.4 and equation (2.4) one derives that sab is a standard word of length π_w. By Lemma 1.2.2, z_w is a conjugate of sab, so that by Theorem 2.1.10 one obtains $w \in St$. ☐

We recall that L_w denotes the minimal integer k for which w has no left special factor of length k. By symmetrical arguments, one can easily prove a result analogous to Proposition 2.1.13, namely, if $\pi_w = L_w + 1$, then $w \in St$.

Examples 2.1.14. The word $w = abbab$ has minimal period $\pi_w = 3$ and $R_w = 2$, therefore it is Sturmian. The word $v = aabba$ is not Sturmian, and indeed $\pi_v = 4 > 3 = R_v + 1 = L_v + 1$. However, for $u = aabab \in St$ one has $\pi_u = 5 > 4 = \max\{R_u, L_u\} + 1$.

Our second characterization of finite Sturmian words is a modification of Proposition 2.1.13:

Theorem 2.1.15. *A finite nonempty word w is Sturmian if and only if*

$$\pi_w = R_{z_w^2} + 1 \ . \tag{2.8}$$

Proof. Assume (2.8) holds. By Lemma 1.2.3, one has $\pi_{z_w^2} = |z_w| = \pi_w = R_{z_w^2} + 1$, so that from Proposition 2.1.13 it follows $z_w^2 \in St$. As $w \in \operatorname{Fact} z_w^*$, one obtains $w \in St$ by Proposition 2.1.8.

Conversely, let $w \in St$. The result is trivial if $\pi_w = 1$, so assume $|z_w| > 1$. By Theorem 2.1.10, z_w is a conjugate of a standard word. Since all conjugates of z_w are factors of z_w^2, by (2.4) and Proposition 2.1.4 there exists $v \in PER$ such that vab and vba are factors of z_w^2, of length π_w. This means that v is a right special factor of z_w^2 of length $\pi_w - 2$; thus $R_{z_w^2} \geq \pi_{z_w^2} - 1$. By (1.1), one has $\pi_{z_w^2} \geq R_{z_w^2} + 1$, hence $\pi_w = \pi_{z_w^2} = R_{z_w^2} + 1$ as desired. □

We remark that in the case of palindromes, condition (2.8) in the preceding theorem can be replaced by the equation $\pi_w = R_w + 1$. This will be proved in Theorem 2.1.37, as a consequence of Proposition 2.1.13 and of a property of Sturmian palindromes (cf. Proposition 2.1.32).

Proposition 2.1.16. *Let w be a word having minimal period $\pi_w > 1$ and v be its shortest prefix such that $\pi_v = \pi_w$. Let ux ($x \in A$) be the suffix of v of length $\pi_w - 1$. One has $w \in St$ if and only if there exists a letter $y \neq x$ such that uy is a factor of z_w^2.*

Proof. If $uy \in \operatorname{Fact} z_w^2$, then u is a right special factor of z_w^2 of length $\pi_w - 2$, so that $\pi_w \leq R_{z_w^2} + 1$. By (1.1) one has $\pi_w = \pi_{z_w^2} \geq R_{z_w^2} + 1$; thus $\pi_w = R_{z_w^2} + 1$ and by Theorem 2.1.15 it follows $w \in St$.

Conversely, as shown in the proof of Theorem 2.1.10, any word of St has an unbordered factor of maximal length, whose value is the minimal period of the word. Therefore, one can write v as $v = tx$ with $x \in \mathcal{A}$ and $\pi_t < \pi_w$ and t cannot have unbordered factors of length π_w since the maximal length of these factors is π_t. Since $v \in St$, it has an unbordered factor r of maximal length $|r| = \pi_v = \pi_w$. This factor has to be necessarily a suffix of v. Since r is unbordered and $|r| = \pi_w > 1$, from Propositions 2.1.5 and 2.1.6 one has $r = yux$ with $u \in PER$ and $\{x, y\} = \mathcal{A}$. By Lemma 1.2.2, z_w is conjugate of yux and, by Proposition 2.1.4, of xuy. Since $xuy \in \operatorname{Fact} z_w^2$, the result follows. $\square$

Examples 2.1.17. Let $w = aababaa \in St$. One has $\pi_w = 5$, $z_w^2 = aababaabab$, and $R_{z_w^2} = 4$, so that $\pi_w = R_{z_w^2} + 1$. The shortest prefix v of w such that $\pi_v = \pi_w$ is $v = aabab$. Its suffix of length $\pi_w - 1$ is $ub = abab$, and $ua = abaa$ is a factor of z_w^2.

Let $r = baabb \notin St$. One has $\pi_r = 4$, $z_r^2 = baabbaab$, and $R_{z_r^2} = 2$, so that $\pi_r > R_{z_r^2} + 1$. In this case, the shortest prefix v such that $\pi_v = \pi_r$ is $v = r$. The suffix ub of v of length 3 is abb, and $aba \notin \operatorname{Fact} z_r^2$.

Enumeration of primitive Sturmian words

As an application of preceding results, we give a formula which counts for any $n > 1$ the finite primitive Sturmian words of length n. We need the following:

Lemma 2.1.18. *The number of words of length $n > 0$ which are conjugate of standard Sturmian words is 2 if $n = 1$ and $n\phi(n)$ for $n > 1$, where ϕ is Euler's totient function.*

Proof. For $n = 1$ the result is trivial since the only two words conjugate of standard words are a and b. Let us suppose $n > 1$. As is well known (see for instance [45, Chap. 2]), the number of standard words of length $n > 1$ is given by $2\phi(n)$. If s is a standard word, by (2.4) we can write $s = vxy$ with $\{x, y\} = \{a, b\}$ and $v \in PER$. By Proposition 2.1.4, $s' = vyx \in Stand$ is a conjugate of s. In the conjugacy class of s there is no other standard word. Indeed, if $t = uxy$ is a conjugate of s, with $u \in PER$, then $|t|_a = |s|_a$ and $|t|_b = |s|_b$, so that t and s have the same "slope"; from this it follows that $u = v$

(see for instance [6, 45]). Hence, in each conjugacy class of a standard word of length $n > 1$ there are exactly two standard words. Thus, the number of these conjugacy classes is $\phi(n)$. Since any standard word is primitive, in any class there are n words. From this the assertion follows. □

Proposition 2.1.19. *For any $n > 1$, the number of primitive finite Sturmian words of length n is given by:*

$$\sum_{i=1}^{n}(n+1-i)\phi(i) - \sum_{\substack{d|n \\ d\neq n}} d\phi(d) \ .$$

Proof. Let w be a non-primitive Sturmian word of length $n > 1$. The word w can be written uniquely as $w = u^k$, with $u \in \pi(\mathcal{A}^*)$ and $k > 1$. Moreover, from Lemma 1.2.3 one has $z_w = u$; by Theorem 2.1.10, u is a conjugate of a standard word. Since $|w| = k|u|$, the integer $|u|$ is a proper divisor of n. Conversely, if u is a conjugate of a standard word, then by Proposition 2.1.8 one has that $u^k \in St$ for any k.

The number of primitive Sturmian words of length n is then obtained by subtracting from $\mathrm{card}(St \cap \mathcal{A}^n)$ the number of words conjugate of a standard word whose length is a proper divisor of n. It is well known (see for instance [45, Chap. 2]) that the number of all finite Sturmian words of length n is given by the following formula:

$$\mathrm{card}(St \cap \mathcal{A}^n) = 1 + \sum_{i=1}^{n}(n+1-i)\phi(i) \ .$$

From Lemma 2.1.18 it follows

$$\mathrm{card}(St \cap \pi(\mathcal{A}^*) \cap \mathcal{A}^n) = 1 + \sum_{i=1}^{n}(n+1-i)\phi(i) - \left(\sum_{\substack{d|n \\ d\neq n}} d\phi(d) + 1 \right)$$

which proves the assertion. □

2.1.3 Sturmian palindromes: structural properties

In the remaining part of this chapter we shall be interested in the set $St \cap PAL$, whose elements will be called *Sturmian palindromes*.

One has that $PER \subseteq St \cap PAL$. However, the previous inclusion is strict since there exist non-central Sturmian palindromes, for instance *abba*.

We have seen that $St = \mathrm{Fact}(PER)$. We shall prove (cf. Corollary 2.1.21) a similar property for Sturmian palindromes.

Theorem 2.1.20. *Every palindromic factor of a standard Sturmian word c_α is a median factor of a palindromic prefix of c_α.*

The result is attributed to A. de Luca [24] by J.-P. Borel and C. Reutenauer, who gave a geometrical proof in [9]. Theorem 2.1.20 can be also obtained as a consequence of a more general result of X. Droubay, J. Justin, and G. Pirillo [29]. We shall report later a direct proof for the sake of completeness.

Corollary 2.1.21. *A word is a Sturmian palindrome if and only if it is a median factor of some central word.*

Proof. Trivially, every median factor of a palindrome is itself a palindrome. Since $St = \mathrm{Fact}(PER)$, it follows that a median factor of an element of PER is a Sturmian palindrome.

Conversely, let u be in $St \cap PAL$. By definition, there exists an infinite (standard) Sturmian word s such that $u \in \mathrm{Fact}\, s$. By Theorem 2.1.20, u is a median factor of a palindromic prefix of s. Since palindromic prefixes of standard Sturmian words are exactly the elements of PER, the result follows. □

Our proof of Theorem 2.1.20, which follows a simple argument suggested by A. Carpi [17], is based on the following results (see [24]):

Proposition 2.1.22. *If $w \in \mathrm{Fact}\, x$, where x is an infinite Sturmian word, then the reversal $\tilde{w}$ is a factor of x too. Moreover, if x is standard, then w is a right special factor of x if and only if $\tilde{w}$ is a prefix of x.*

Corollary 2.1.23. *A palindromic factor of an infinite standard Sturmian word x is a right special factor of x if and only if it is a palindromic prefix of x.*

Proof of Theorem 2.1.20. By contradiction, let $c_\alpha = \lambda u x$, where u is a palindrome that is not a median factor of any palindromic prefix of c_α, and $\lambda \in \mathcal{A}^*$ has minimal length for such condition. Since u cannot be a prefix of c_α, we have $|\lambda| \geq 1$. Thus we can assume, without loss of generality, $\lambda = \lambda' a$. Now let

z be the first letter of x, so that $x = zx'$. Suppose first $z = a$. The palindrome aua is not a median factor of a palindromic prefix of c_α, otherwise so would be u. But $c_\alpha = \lambda' auax'$ with $|\lambda'| < |\lambda|$, and this contradicts the minimality of $|\lambda|$. Therefore $z = b$, and then aub and $bua = \widetilde{aub}$ are factors of c_α, in view of Proposition 2.1.22. This means in particular that u is a right special factor of c_α. Corollary 2.1.23 then implies that u is a prefix of c_α, a contradiction. □

We recall some basic facts (see [28, 24]):

Proposition 2.1.24. *Let w be a word. The following conditions are equivalent:*

1. *$w \in PER$,*
2. *awb and bwa are Sturmian,*
3. *awa, awb, bwa, and bwb are all Sturmian.*

Proposition 2.1.25. *If wa and wb are Sturmian words, then there exists a letter $x \in \mathcal{A}$ such that xwa and xwb are both Sturmian.*

We now prove two easy consequences (see also [24]):

Proposition 2.1.26. *Let $w \in \mathcal{A}^*$ be a palindrome. If wa and wb are Sturmian, then w is central.*

Proof. From the previous proposition, there exists a letter $x \in \mathcal{A}$ such that xwa and xwb are both Sturmian. Without loss of generality, we may suppose $x = a$, so that $awb \in St$. Therefore $\widetilde{awb} = bwa$ is Sturmian too, thus by Proposition 2.1.24, w is central. □

Lemma 2.1.27. *Let w be a Sturmian palindrome. If w is not central, then there exists a unique letter $x \in \mathcal{A}$ such that xwx is Sturmian.*

Proof. If awa and bwb are both Sturmian, then $w \in PER$ by Proposition 2.1.26, a contradiction. Since by Corollary 2.1.21 the word w is a (proper) median factor of some central word, there exists a unique letter $x \in \mathcal{A}$ such that xwx is Sturmian. □

We have seen with Corollary 2.1.21 that a Sturmian palindrome is a median factor of a central word. We will now give some further results concerning the structure of Sturmian palindromes.

Proposition 2.1.28. *A palindrome $w \in A^*$ with minimal period $\pi_w > 1$ can be uniquely represented as*

$$w = w_1 x y w_2 = w_2 y x \tilde{w}_1$$

with $x, y \in A$, w_2 the longest proper palindromic suffix of w, and $|w_1 xy| = \pi_w$. The word w is not central if and only if either $w_1 \notin PAL$ or $w = (w_1 xx)^k w_1$ where $k \geq 1$ is the order of w.

Proof. Since $\pi_w > 1$, it follows by Lemma 3.1.3 that w can be uniquely factorized as $w = w_1 x y w_2$ where w_2 is the longest proper palindromic suffix of w, $x, y \in A$, and $|w_1 xy| = \pi_w$. Since w is a palindrome, we can write

$$w = w_1 x y w_2 = w_2 y x \tilde{w}_1 \ .$$

When $\pi_w > 1$, by Proposition 2.1.1, w is central if and only if $w_1 \in PAL$ and $x \neq y$. Therefore, in the case $w_1 \in PAL$, w is not central if and only if $w = w_1 xx w_2 = w_2 xx w_1$. The word w has the two periods

$$\pi_w = |w_1 xx| \ \text{ and } q = |w_2 xx| \tag{2.9}$$

and length $\pi_w + q - 2$. Thus $w \notin PER$ if and only if $d = \gcd(\pi_w, q) > 1$. Since $|w| \geq \pi_w + q - d$, by Theorem 1.2.1 the word w has the period $d = \pi_w$. This occurs if and only if $q = k\pi_w$ with $k \geq 1$. From (2.9) this condition is equivalent to the statement $w_2 xx = (w_1 xx)^k$, i.e., $w = (w_1 xx)^k w_1$. □

Example 2.1.29. Let $w = aaabaaaaaabaaa \in St \cap PAL$, with $\pi_w = 7$. The word w can be factorized as $(aaaba)aa(aaabaaa)$, where $aaabaaa$ is the longest proper palindromic suffix of w, $|aaaba| = \pi_w - 2 = 5$. The prefix $aaaba$ is not a palindrome, thus w is not central.

Let $v = abaababababaaba \in St \cap PAL$. We factorize v as

$$v = (abaabab)ab(abaaba)$$

where $abaaba$ is the longest proper palindromic suffix of v. Also in this case $abaabab$ is not a palindrome, so that $w \notin PER$.

Let $u = abbabbabba \in St \cap PAL$. We factorize u as $(a)bb(abbabba)$, where $abbabba$ is the longest palindromic suffix of u. In this case, the prefix a is a palindrome, and $u = (abb)^3a$. Hence u is not central.

Lemma 2.1.30. *If $w = w_1xyw_2 = w_2yx\tilde{w}_1$, where w_2 is the longest proper palindromic suffix of w and $x, y \in A$, then $w' = ywy$ has the minimal period $\pi_{w'} = \pi_w$.*

Proof. Since w is a factor of w', one has $\pi_{w'} \geq \pi_w$. The word yw_2y is a palindromic proper suffix of $w' = yw_1xyw_2y$, so that by Lemma 3.1.3 the word w' has the period $|yw_1x|$. Hence, $\pi_{w'} \leq |yw_1x| = |w_1xy| = \pi_w$. Thus $\pi_w = \pi_{w'}$. □

The next lemma is essentially a restatement of Lemma 2 in [22]. Note that its first part is an obvious consequence of Lemma 2.1.30.

Lemma 2.1.31. *Let $w = w_1xyw_2 = w_2yxw_1 \in PER$, with $|w_2| > |w_1|$ and $\{x, y\} = \mathcal{A}$. The word $v = ywy$ has minimal period $\pi_v = \pi_w$, whereas $v' = xwx = xw_1xyw_2x$ has minimal period $\pi_{v'} = |w_2| + 2 = |w| - \pi_w + 2$.*

Let $w \in (St \cap PAL) \setminus PER$. We denote by u the (unique) *shortest* median extension of w in PER, and by v the *longest* central median factor of w. Note that also v is unique. For instance, for the Sturmian palindrome $w = baaabaaab$ one has $u = aawaa$ and $v = aaabaaa$.

Theorem 2.1.32. *Let $w \in (St \cap PAL) \setminus PER$. With the preceding notation, one has $\pi_u = \pi_w$. Moreover, either $\pi_w = \pi_v$ or $\pi_w = |v| - \pi_v + 2$.*

Proof. We consider first the case that $\pi_v = 1$, so that $v = x^n$ with $x \in \mathcal{A}$ and $n = |v|$. In such a case w has also the median palindromic factor $v_1 = yx^ny$, where $\{x, y\} = \mathcal{A}$ (recall that v is the longest central median factor of w). Moreover, $n = |v|$ is at least 2, otherwise v_1 would be equal to $yxy \in PER$. One has $\pi_{v_1} = |yx^n| = n + 1 = |v| - \pi_v + 2$. Now we define, for $2 \leq i \leq n$:

$$v_i = xv_{i-1}x = x^{i-1}yx^nyx^{i-1} = (x^{i-1}yx^{n-i+1})(x^{i-1}yx^{i-1}) . \tag{2.10}$$

The word $v_n = x^{n-1}yx^nyx^{n-1}$ is central, whereas by Lemma 2.1.3 we have $v_i \notin PER$. From Lemma 2.1.27 it follows that the words v_i are the *only*

Sturmian extensions of v_1 which are median factors of v_n. Since for $i < n$ one has $v_i \notin PER$, one derives that $w = v_k$ for some $1 \le k < n$, and $u = v_n$. As shown in (2.10), by Lemma 2.1.30 all the v_i's have the same minimal period, for $1 \le i \le n$. The result in this case follows: $\pi_w = \pi_u = |v| - \pi_v + 2$.

Now let us assume $\pi_v > 1$. One has $v = w_1xyw_2 = w_2yxw_1$, with $w_1, w_2 \in PAL$ and $x \neq y$. We suppose $|w_1| < |w_2|$, so that $\pi_v = |w_1| + 2$. From the definition of v, it follows that there exists a letter $z \in \mathcal{A}$ such that $v_1 = zvz$ is a median factor of w which is not central. By Lemma 2.1.31, we have $\pi_{v_1} = \pi_v$ if $z = y$, or else $\pi_{v_1} = |v| - \pi_v + 2$ if $z = x$.

Using Lemma 2.1.30, we shall now define a sequence of palindromes with the same minimal period as v_1. Let us first suppose that $z = y$, so that $v_1 = yw_1xyw_2y$. We set $v_2 = xv_1x = (xyw_1)(xyw_2yx)$. Moreover, if $w_1 = p_1p_2 \cdots p_k$ with $p_j \in \mathcal{A}$ for $1 \le j \le k$, we set $v_i = p_{k-i+3}v_{i-1}p_{k-i+3}$ for $i \ge 3$, so that

$$\begin{aligned} v_3 &= p_kv_2p_k = (p_kxyp_1 \cdots p_{k-1})(p_kxyw_2yxp_k)\ , \\ &\vdots \\ v_{k+2} &= p_1v_{k+1}p_1 = p_1 \cdots p_kxyw_1xyw_2yxp_k \cdots p_1 = w_1xyw_1xyw_2yx\tilde{w}_1\ . \end{aligned}$$

Since $w_1 = \tilde{w}_1$, the last equation can be written as

$$v_{k+2} = (w_1)xy(w_1xyw_2yxw_1) = (w_1xyw_2yxw_1)yx(w_1)$$

showing, by Proposition 2.1.28, that the word v_{k+2} is central, so that for any $i \le \pi_v = k + 2$ one has $v_i \in St \cap PAL$.

Let $s \le k + 2$ be the minimal integer such that $v_s \in PER$. Since for $i < s$ one has $v_i \notin PER$, one derives from Lemma 2.1.27 that $u = v_s$ and $w = v_r$ for some integer $r < s$. Hence $\pi_w = \pi_{v_s} = \pi_u$, and in this case $\pi_w = \pi_v$.

The case $z = x$ is similarly dealt with, but interchanging the roles of w_1 and w_2. Thus one assumes $w_2 = q_1 \cdots q_k$, $q_j \in \mathcal{A}$, $1 \le j \le k$, and defines v_i as $q_{k-i+3}v_{i-1}q_{k-i+3}$ for $i \ge 3$, starting from $v_2 = yv_1y = (yxw_2)(yxw_1xy)$ and ending with

$$v_{k+2} = w_2yxw_2yxw_1xyw_2 \in PER\ .$$

Therefore there exist integers r and s such that $1 \le r < s \le k+2 = |v| - \pi_v + 2$, $w = v_r$, and $u = v_s$, so that $\pi_w = \pi_u$ and $\pi_w = \pi_{v_1} = |v| - \pi_v + 2$. $\square$

Example 2.1.33. Let $w = baaabaaab \in St \cap PAL$. Following the notations of Theorem 2.1.32, one has $v = aaabaaa$, $v_1 = w$, and $u = v_3 = aabaaabaaabaa$. Thus $\pi_w = \pi_u = \pi_v = 4$.

Let $w = babbbbab$. In this case we have $v = bbbb$, $w = v_2$, and $u = v_4 = bbbabbbbabbb$, so that $\pi_w = \pi_u = 5 = |v| + 1 = |v| - \pi_v + 2$.

For any word $w \in A^*$, we denote by K_w the length of the shortest unrepeated suffix of w. Conventionally, one assumes $K_\varepsilon = 0$. There exist some relations among the parameters R_w, K_w, π_w, and $|w|$; the following lemma synthesizes some results proved in [25, Corollary 5.3, Propositions 4.6 and 4.7] which will be useful in the sequel.

Lemma 2.1.34. *For any $w \in A^*$, one has*

$$|w| \geq R_w + K_w \ .$$

Moreover, the following holds:

- *if $\pi_w = R_w + 1$, then $|w| = R_w + K_w$,*
- *if $|w| = R_w + K_w$, then for any n there exists at most one right special factor of w of length n.*

The following theorem gives a further criterion, different from Proposition 2.1.28, to discriminate whether a palindrome is central or not.

Theorem 2.1.35. *Let $w \in A^*$ be a palindrome with $\pi_w > 1$. Then w is central if and only if its prefix of length $\pi_w - 2$ is a right special factor of w.*

Proof. From Proposition 2.1.28, we can write

$$w = w_1 xy w_2 = w_2 yx \tilde{w}_1 \tag{2.11}$$

where $x, y \in A$, w_2 is the longest proper palindromic suffix of w, $|w_1| = \pi_w - 2$, and w is central if and only if $w_1 \in PAL$ and $x \neq y$. Therefore we have to prove that w_1 is a right special factor of w if and only if $w_1 = \tilde{w}_1$ and $x \neq y$.

Indeed, assume that these two latter conditions are satisfied. Since $\tilde{w}_1 = w_1$ and w_2 is the longest proper palindromic suffix (and prefix) of w, one has that

w_1 is a border of w_2. This implies, from (2.11), that w_1 is a right special factor of w.

Conversely, suppose w_1 is a right special factor of w. Let us first prove that $w_1 \in PAL$. By hypothesis, we have $\pi_w - 2 = |w_1| \leq R_w - 1$, that is $R_w \geq \pi_w - 1$. By Lemma 2.1.34 one has $\pi_w \geq R_w + 1$, so that $\pi_w = R_w + 1$. This implies $|w| = R_w + K_w$, again by Lemma 2.1.34. The suffix $\tilde{w}_1$ of w is repeated, because w_1 is a right special factor of w, which is a palindrome. This leads to

$$\pi_w - 2 = |\tilde{w}_1| \leq K_w - 1$$

and thus to $|w| = R_w + K_w \geq 2\pi_w - 2$. If $|w| = 2\pi_w - 2$, then $|w_1| = |w_2|$ so that one derives $w_1 = w_2 \in PAL$. If $|w| \geq 2\pi_w - 1$, then w has the prefix $w_1 x y w_1 x$, so that $y w_1 x \in \operatorname{Fact} w$. Since w_1 is a right special factor of w, there exists a letter $z \neq x$ such that $w_1 z \in \operatorname{Fact} w$. Moreover, since $w_1 z$ is not a prefix, there exists a letter y' such that $y' w_1 z \in \operatorname{Fact} w$. One has $y \neq y'$, for otherwise $y w_1$ would be a right special factor of w of length $\pi_w - 1 = R_w$, which is a contradiction. As w is a palindrome, the words $x \tilde{w}_1 y$ and $z \tilde{w}_1 y'$ are factors of w too, so that $\tilde{w}_1$ is a right special factor of w. By Lemma 2.1.34, one obtains $w_1 = \tilde{w}_1$. Therefore we get $w_1 \in PAL$ again.

We shall now prove that $x \neq y$. By contradiction, suppose w has the factorization

$$w = (w_1 x x)^k w_1 \,, \text{ with } k \geq 1$$

as granted by Proposition 2.1.28. Since w_1 is a right special factor of w, one has $w_1 z \in \operatorname{Fact} w$ for a suitable letter $z \neq x$. Thus we have either $w_1 z = x w_1$ or $w_1 z = v_2 x x v_1 z$, where $v_1 z$ is a prefix of w_1 and v_2 is a suffix of w_1. Since $|w_1| = |w_1 z| - 1$, we can write $w_1 = v_1 z \alpha v_2$, with $\alpha \in A$. The first case is impossible since w_1 is a palindrome and $x \neq z$. In the latter case, one obtains:

$$v_1 z \alpha v_2 = w_1 = \tilde{w}_1 = \tilde{v}_1 x x \tilde{v}_2$$

which is absurd again, because $x \neq z$. □

Example 2.1.36. The word $w = baab$ is a Sturmian palindrome of minimal period $\pi_w = 3$. Its prefix of length 1 is not a right special factor, hence $w \notin PER$. The word $v = abababbababa$ is a Sturmian palindrome having minimal period 7, and its prefix $ababa$ of length 5 is not right special. Therefore

$v \notin PER$. On the contrary, the word $u = aabaabaa$ has minimal period 3, and its prefix of length 1 is a right special factor, so that u is central.

In Proposition 2.1.13 we have proved that any finite word w such that $\pi_w = R_w + 1$ is Sturmian. The converse does not hold in general, as shown in Examples 2.1.14 and 2.1.38. However, the result is true for Sturmian palindromes, as the next theorem shows.

Theorem 2.1.37. *A palindrome $w \in A^*$ is Sturmian if and only if $\pi_w = R_w + 1$.*

Proof. By Proposition 2.1.13, the condition is sufficient. Necessity is trivially true if $\pi_w = 1$. By (1.1), one has $\pi_w \geq R_w + 1$. Hence, if $\pi_w > 1$ the condition $\pi_w = R_w + 1$ is equivalent to the existence of a right special factor s of w of length $|s| = \pi_w - 2$.

We prove that every Sturmian palindrome w such that $\pi_w \geq 2$ has such a factor. If w is central, the result follows directly from Theorem 2.1.35. Thus we suppose $w \notin PER$, and as in Theorem 2.1.32 we denote by v the central median factor of w of maximal length.

If $\pi_v = 1$, then there exists a letter $x \in \mathcal{A}$ and an integer $n \geq 1$ such that $v = x^n$. From the maximality condition, one derives that $n > 1$. In this case, by Theorem 2.1.32 one derives $\pi_w = |v| + 1 = n + 1$ and $yx^ny \in \operatorname{Fact} w$, where $\{x, y\} = \mathcal{A}$; therefore x^{n-1} is the desired right special factor of w, of length $n - 1 = \pi_w - 2$.

If $\pi_v > 1$, by using Proposition 2.1.1 we can write v as $v_1xyv_2 = v_2yxv_1$, with $\pi_v = |v_1xy|$. By Theorem 2.1.32, one has either $\pi_w = \pi_v$ or $\pi_w = |v| - \pi_v + 2$. In the first case, the result is a consequence of Theorem 2.1.35. Indeed, the prefix v_1 of the central word v, whose length is $\pi_v - 2 = \pi_w - 2$, is a right special factor of v, and then of w. In the latter case, one derives that the word $xvx = xv_1xyv_2x = xv_2yxv_1x$ is a factor of w, so that v_2 is a right special factor of w, of length $|v| - \pi_v = \pi_w - 2$. □

Example 2.1.38. The word $u = ababaa$ is not a palindrome, but $\pi_u = 5 = R_u + 1$, thus it is Sturmian. However, the word $v = aabab \in St$ has $\pi_v = 5 > 3 = R_v + 1$. Let $w = abba \in St \cap PAL$. One has $\pi_w = 3 = R_w + 1$. The palindrome $s = aabbaa$ is not Sturmian. One has $\pi_s = 4 > 3 = R_s + 1$.

2.2 Episturmian words

A word $w \in A^\omega$ is called *Episturmian* if it is closed under reversal and it has at most one right (or equivalently, left) special factor of each length. We recall (see [29]) that every Episturmian word is *uniformly recurrent*, i. e., every factor of an Episturmian word occurs infinitely often, with bounded gaps.

An Episturmian word w is called *standard* if every left special factor of w is a prefix of it. We denote by $Ep(A)$, or simply Ep, the set of all Episturmian words over A, and by SEp the set of standard ones.

Proposition 2.2.1 (cf. [29]). *For every Episturmian word w, there exists a standard Episturmian word s such that* $\mathrm{Fact}(s) = \mathrm{Fact}(w)$.

Thus $\mathrm{Fact}(Ep) = \mathrm{Fact}(SEp)$. The elements of $\mathrm{Fact}(Ep)$ are called *finite Episturmian words.*

Given a word $w \in A^*$, we denote by $w^{(+)}$ its *right palindrome closure*, i. e., the shortest palindrome having w as a prefix. Similarly, $w^{(-)}$ is the left palindrome closure of w. For instance, if $w = abacbca$, then $w^{(+)} = abacbcaba$ and $w^{(-)} = acbcabacbca$.

For any $w \in A^*$, one has $w^{(-)} = \tilde{w}^{(+)}$. Moreover, if Q is the longest palindromic suffix of w and $w = sQ$, then $w^{(+)} = sQ\tilde{s}$.

Let $\psi : A^* \to A^*$ be defined by $\psi(\varepsilon) = \varepsilon$ and $\psi(va) = (\psi(v)a)^{(+)}$ for any $a \in A$ and $v \in A^*$. For any $u, v \in A^*$, one has $\psi(uv) \in \psi(u)A^* \cap A^*\psi(u)$. The map ψ can then be naturally extended to A^ω by setting, for any infinite word x,

$$\psi(x) = \lim_{n\to\infty} \psi(w_n) \ ,$$

where $\{w_n\} = \mathrm{Pref}(x) \cap A^n$ for all $n \geq 0$.

Proposition 2.2.2 (cf. [29]). *Let $s \in A^\omega$. The following conditions are equivalent:*

1. *s is a standard Episturmian word,*
2. *for any prefix u of s, $u^{(+)}$ is also a prefix of s,*
3. *there exists $x \in A^\omega$ such that $s = \psi(x)$.*

Given a standard Episturmian word s, the (unique) infinite word x such that $s = \psi(x)$ is called *directive word* of s and is denoted by $\Delta(s)$, or simply by Δ. From the preceding proposition, one can easily derive (cf. [29]) that the set of palindromic prefixes of a standard Episturmian word s coincides with

$$\{\psi(u) \mid u \in \mathrm{Pref}(\Delta(s))\} .$$

A standard Episturmian word s over the alphabet A is called a (standard) *Arnoux-Rauzy* word if every symbol of A occurs infinitely often in the associated directive word $\Delta(s)$. We will denote by $AR(A)$, or simply AR, the set of Arnoux-Rauzy words over A. In the case of a binary alphabet, an AR-word is usually called *standard Sturmian word.*

Example 2.2.3. Let $A = \{a, b\}$ and $x = (ab)^\omega$. One has that

$$f = \psi(x) = abaababaabaababa \cdots$$

is the famous *Fibonacci word*, a standard Sturmian word. On an alphabet with three letters $A = \{a, b, c\}$, if we take $x = (abc)^\omega$ as a directive word, then

$$\tau = \psi(x) = abacabaabacababacabaabac \cdots$$

is a standard Arnoux-Rauzy word, often called *Tribonacci* word. The word $s = cabaabacababacabaab \cdots$ such that $abas = \tau$ is an example of an Episturmian word which is not standard, as a is a left special factor of s but not a prefix of it.

The periodic word $s = (abac)^\omega$ is standard Episturmian, but not Arnoux-Rauzy. Its directive word is $\Delta(s) = abc^\omega$.

The following proposition can be easily proved using well-known results on Episturmian words (see [29]).

Proposition 2.2.4. *Let s be a standard Episturmian word. Any bispecial factor of s is a palindromic prefix of s. If s is not periodic, the converse holds too.*

Proposition 2.2.5. $\mathrm{Fact}(Ep) = \mathrm{Fact}(AR)$.

Proof. Let $u \in \mathrm{Fact}(Ep) = \mathrm{Fact}(SEp)$. Hence there exists $s \in SEp$ such that $u \in \mathrm{Fact}(s)$. Now let be $s = \psi(\Delta)$ where $\Delta = t_1 t_2 \cdots t_n \cdots$, with $t_i \in A$ for $i \geq 1$. Therefore there exists a palindromic prefix p of s such that $u \in \mathrm{Fact}(p)$. Now $p = \psi(t_1 \cdots t_i)$ for some i. We can consider $\Delta' = t_1 \cdots t_i t$ with $t \in A^\omega$ such that any letter of A occurs infinitely many times in t. Hence $s' = \psi(\Delta') \in AR$ and contains p as a factor, so that $u \in \mathrm{Fact}(s')$. Therefore, $\mathrm{Fact}(Ep) \subseteq \mathrm{Fact}(AR)$. Since the inverse inclusion is trivial, the result follows. $\square$

The following proposition collects two properties of standard Episturmian words (cf. Lemmas 1 and 4 in [29]) which will be useful in the sequel.

Proposition 2.2.6 (cf. [29]). *Let s be a standard Episturmian word. The following hold:*

1. *Any prefix p of s has a palindromic suffix which has a unique occurrence in p.*
2. *The first letter of s occurs in every factor of s having length 2.*

Clearly, if p is a prefix of a standard Episturmian word, then the palindromic suffix of p which has a unique occurrence in p is the longest palindromic suffix of p.

We recall (cf. [29, 43]) that an infinite word $t \in A^\omega$ is *standard Episturmian* if it is *closed under reversal* (that is, if $w \in \mathrm{Fact}\, t$ then $\tilde{w} \in \mathrm{Fact}\, t$) and each of its left special factors is a prefix of t. We denote by $SEpi(A)$, or by $SEpi$ when there is no ambiguity, the set of all standard Episturmian words over the alphabet A.

Given a word $w \in A^*$, we denote by $w^{(+)}$ its *right palindrome closure*, i.e., the shortest palindrome having w as a prefix (cf. [24]). If Q is the longest palindromic suffix of w and $w = sQ$, then $w^{(+)} = sQ\tilde{s}$. For instance, if $w = abacbca$, then $w^{(+)} = abacbcaba$.

We define the *iterated palindrome closure* operator $\psi : A^* \to A^*$ by setting $\psi(\varepsilon) = \varepsilon$ and $\psi(va) = (\psi(v)a)^{(+)}$ for any $a \in A$ and $v \in A^*$. From the definition, one easily obtains that the map ψ is injective. Moreover, for any $u, v \in A^*$, one has $\psi(uv) \in \psi(u)A^* \cap A^*\psi(u)$. The operator ψ can then be naturally extended to A^ω by setting, for any infinite word x,

$$\psi(x) = \lim_{n \to \infty} \psi(x_{[n]}) \ .$$

The following fundamental result was proved in [29]:

Theorem 2.2.7. *An infinite word t is standard Episturmian over A if and only if there exists $\Delta \in A^\omega$ such that $t = \psi(\Delta)$.*

For any $t \in SEpi$, there exists a *unique* Δ such that $t = \psi(\Delta)$. This Δ is called the *directive word* of t. If every letter of A occurs infinitely often in Δ, the word t is called a (standard) *Arnoux-Rauzy word.* In the case of a binary alphabet, an Arnoux-Rauzy word is usually called a *standard Sturmian word* (cf. [8]).

Example 2.2.8. Let $A = \{a, b\}$ and $\Delta = (ab)^\omega$. The word $\psi(\Delta)$ is the famous *Fibonacci word*

$$\mathrm{f} = abaababaabaababaababa \cdots .$$

If $A = \{a, b, c\}$ and $\Delta = (abc)^\omega$, then $\psi(\Delta)$ is the so-called *Tribonacci word*

$$\tau = abacabaabacababacabaabacabaca \cdots .$$

A letter $a \in A$ is said to be *separating* for $w \in A^\infty$ if it occurs in each factor of w of length 2. We recall the following well known result from [29]:

Proposition 2.2.9. *Let t be a standard Episturmian word and a be its first letter. Then a is separating for t.*

For instance, the letter a is separating for f and τ.

We report here some properties of the operator ψ which will be useful in the sequel. The first one is known (see for instance [24, 29]); we give a proof for the sake of completeness.

Proposition 2.2.10. *For all $u, v \in A^*$, u is a prefix of v if and only if $\psi(u)$ is a prefix of $\psi(v)$.*

Proof. If u is a prefix of v, from the definition of the operator ψ, one has that $\psi(v) \in \psi(u)A^* \cap A^*\psi(u)$, so that $\psi(u)$ is a prefix (and a suffix) of $\psi(v)$. Let us now suppose that $\psi(u)$ is a prefix of $\psi(v)$. If $\psi(u) = \psi(v)$, then, since ψ is injective, one has $u = v$. Hence, suppose that $\psi(u)$ is a proper prefix of $\psi(v)$. If $u = \varepsilon$, the result is trivial. Hence we can suppose that $u, v \in A^+$. Let $v = a_1 \cdots a_n$ and i be the integer such that $1 \leq i \leq n-1$ and

$$|\psi(a_1 \cdots a_i)| \leq |\psi(u)| < |\psi(a_1 \cdots a_{i+1})|.$$

If $|\psi(a_1 \cdots a_i)| < |\psi(u)|$, then $\psi(a_1 \cdots a_i)a_{i+1}$ is a prefix of the palindrome $\psi(u)$, so that one would have:

$$|\psi(a_1 \cdots a_{i+1})| = |(\psi(a_1 \cdots a_i)a_{i+1})^{(+)}| \leq |\psi(u)| < |\psi(a_1 \cdots a_{i+1})|$$

which is a contradiction. Therefore $|\psi(a_1 \cdots a_i)| = |\psi(u)|$, that implies $\psi(a_1 \cdots a_i) = \psi(u)$ and $u = a_1 \cdots a_i$. $\square$

Proposition 2.2.11. *Let $x \in A \cup \{\varepsilon\}$, $w' \in A^*$, and $w \in w'A^*$. Then $\psi(w'x)$ is a factor of $\psi(wx)$.*

Proof. By the previous proposition, $\psi(w')$ is a prefix of $\psi(w)$. This solves the case $x = \varepsilon$. For $x \in A$, we prove the result by induction on $n = |w| - |w'|$.

The assertion is trivial for $n = 0$. Let then $n \geq 1$ and write $w = ua$ with $a \in A$ and $u \in A^*$. As $w' \in \operatorname{Pref} u$ and $|u| - |w'| = n - 1$, we can assume by induction that $\psi(w'x)$ is a factor of $\psi(ux)$. Hence it suffices to show that $\psi(ux) \in \operatorname{Fact} \psi(wx)$. We can write

$$\psi(w) = (\psi(u)a)^{(+)} = \psi(u)av = \tilde{v}a\psi(u)$$

for some $v \in A^*$, so that $\psi(wx) = (\tilde{v}a\psi(u)x)^{(+)}$. Since $\psi(u)$ is the longest proper palindromic prefix and suffix of $\psi(w)$, if $x \neq a$ it follows that the longest palindromic suffixes of $\psi(u)x$ and $\psi(w)x$ must coincide, so that $\psi(ux) = (\psi(u)x)^{(+)}$ is a factor of $\psi(wx)$, as desired.

If $x = a$, then $\psi(ux) = \psi(w)$ is trivially a factor of $\psi(wx)$. This concludes the proof. $\square$

The following proposition was proved in [29, Theorem 6].

Proposition 2.2.12. *Let $x \in A$, $u \in A^*$, and $\Delta \in A^\omega$. Then $\psi(u)x$ is a factor of $\psi(u\Delta)$ if and only if x occurs in Δ.*

For each $a \in A$, let $\mu_a : A^* \to A^*$ be the morphism defined by $\mu_a(a) = a$ and $\mu_a(b) = ab$ for all $b \in A \setminus \{a\}$. If $a_1, \ldots, a_n \in A$, we set $\mu_w = \mu_{a_1} \circ \cdots \circ \mu_{a_n}$ (in particular, $\mu_\varepsilon = \mathrm{id}_A$). The next proposition, proved in [42], shows a connection between these morphisms and iterated palindrome closure.

Proposition 2.2.13. *For any $w, v \in A^*$, $\psi(wv) = \mu_w(\psi(v))\psi(w)$.*

By the preceding proposition, if $v \in A^\omega$ then one has

$$\begin{aligned}\psi(wv) &= \lim_{n\to\infty} \psi(wv_{[n]}) = \lim_{n\to\infty} \mu_w(\psi(v_{[n]}))\psi(w) \\ &= \lim_{n\to\infty} \mu_w(\psi(v_{[n]})) = \mu_w(\psi(v)) \ .\end{aligned}$$

Thus, for any $w \in A^*$ and $v \in A^\omega$ we have

$$\psi(wv) = \mu_w(\psi(v)) \ . \tag{2.12}$$

Corollary 2.2.14. *For any $t \in A^\omega$ and $w \in A^*$, $\psi(w)$ is a prefix of $\mu_w(t)$.*

Proof. Let $t = t_1 t_2 \cdots t_n \cdots$, with $t_i \in A$ for $i \geq 1$. We prove that $\psi(w)$ is a prefix of $\mu_w(t_{[n]})$ for all n such that $|\mu_w(t_{[n]})| \geq |\psi(w)|$. Indeed, by Proposition 2.2.13 we have, for all $i \geq 1$, $\mu_w(t_i)\psi(w) = \psi(wt_i) = \psi(w)\xi_i$ for some $\xi_i \in A^*$. Hence

$$\mu_w\left(t_{[n]}\right)\psi(w) = \mu_w(t_1)\cdots\mu_w(t_n)\psi(w) = \psi(w)\xi_1\cdots\xi_n \ ,$$

and this shows that $\psi(w)$ is a prefix of $\mu_w(t_{[n]})$. $\square$

From the definition of the morphism μ_a, $a \in A$, it is easy to prove the following:

Proposition 2.2.15. *Let $w \in A^\infty$ and $a = w^f$. Then a is separating for w if and only if there exists $\alpha \in A^\infty$ such that $w = \mu_a(\alpha)$.*

For instance, the letter a is separating for the word $w = abacaaacaba$, and one has $w = \mu_a(bcaacba)$.

2.2.1 A closure property

We want to show that if $w \in \mathrm{Fact}(Ep)$, then also its right and left palindrome closures belong to $\mathrm{Fact}(Ep)$; since Episturmian words are closed under reversal, and $w^{(-)} = \tilde{w}^{(+)}$, it suffices to prove only the right palindrome closure case. We have the following

Proposition 2.2.16. *Let u be a non-palindromic finite Episturmian word; let Q be the longest palindromic suffix of u and write $u = saQ$ where $a \in A$ and $s \in A^*$ (s possibly empty). Then $ua = saQa$ is a finite Episturmian word.*

Before proving the proposition we need some lemmas. The first lemma was proved in [3, Theorem 1.1]. We report here a different and simpler proof.

Lemma 2.2.17. *Let w be an Episturmian word and $P \in PAL \cap \mathrm{Fact}(w)$. Then every first return to P in w is a palindrome.*

Proof. By Proposition 2.2.1, we may always suppose that w is a standard Episturmian word. Let $u \in \mathrm{Fact}(w)$ be a first return to the palindrome P, i. e., $u = P\lambda = \rho P$, $\lambda, \rho \in A^*$, and the only two occurrences of P in u are as a prefix and as a suffix of u. If $|P| > |\rho|$, then the prefix P of u overlaps with the suffix P in u and this implies, as is easily to verify, that u is a palindrome. Then let us suppose that $u = PvP$ with $v \in A^*$.

Now we consider the first occurrence of u or of $\tilde{u}$ in w. Without loss of generality, we may suppose that $w = \alpha u w'$ and that $\tilde{u}$ does not occur in the prefix of w having length $|\alpha u| - 1$. Let Q be the palindromic suffix of αu of maximal length. If $|Q| > |u|$, then we have that $\tilde{u}$ occurs in αu before u, which is absurd. Then suppose $|Q| \leq |u|$. If $|u| > |Q| > |P|$, then one contradicts the hypothesis that u is a first return to P. If $|Q| = |P|$, then $Q = P$ has more than one occurrence in αu, which is absurd in view of Proposition 2.2.6. The only remaining possibility is $Q = u$, i. e., u is a palindrome. □

The following lemma is well known. We report here a proof for the sake of completeness.

Lemma 2.2.18. *Let $w \in AR$ and s be the unique right special factor of length n. If $B_1, \ldots, B_m, \ldots$ are the bispecial factors of w ordered by increasing length, then s is a suffix of any B_m such that $|s| \leq |B_m|$ and, for any $x \in A$, $sx \in \mathrm{Fact}(w)$.*

Proof. Since w is not periodic, by Proposition 2.2.4 the bispecial factors B_i, $i > 0$, are its palindromic prefixes. Moreover, if $t = t_1 t_2 \cdots t_n \cdots \in A^\omega$ is the directive word of w, then $B_{i+1} = (B_i t_i)^{(+)}$ for any $i > 0$. Since s is a right special factor of w, $\tilde{s}$ is left special and thus a prefix of w. Therefore, s is a suffix of any palidromic prefix B_m of w such that $|s| \leq |B_m|$. As $w \in AR$, any letter $x \in A$ occurs infinitely often in t; hence there exists $k \geq m$ such that $x = t_k$, so that $B_k x$ is a factor of w. Since B_m is a suffix of B_k, it follows $sx \in \mathrm{Fact}(w)$. □

Lemma 2.2.19. *Let w and w' be Arnoux-Rauzy words on the alphabet A. If w and w' have the same right special factor of length n, then they share the same factors up to length $n+1$.*

Proof. Trivial if $n = 0$. By induction, suppose we have proved the assertion for the integer $n-1 \geq 0$. Let Q be the common right special factor of w and w' of length n. If we write $Q = aQ'$, with $a \in A$, then Q' is the only right special factor of length $n-1$ of both w and w'. Hence w and w' have the same factors up to length n.

By symmetry, it suffices to prove that any factor v of w, of length $|v| = n+1$, is also a factor of w'. Let $v = v'b$, $b \in A$. Suppose first that $v' = Q$. By Lemma 2.2.18, each right extension Qx, with $x \in A$, is a factor of both w and w'; in particular, v is a factor of w'.

Now assume that $v' \neq Q$. Let $v' = cv''$ with $c \in A$, and suppose that $v'' = Q'$. One has then $c \neq a$. In this case, since $v = cv''b$ and $Qb = av''b$ are different factors of w, one has that $v''b$ is left special in w. Since $|v''b| = n$, one derives that $v''b = \tilde{Q}$ is a left special factor of w' too, so that $v = cv''b$ is a factor of w' as a consequence of Lemma 2.2.18.

If $v'' \neq Q'$, then $v''b$ is the unique right extension of v'' in w. As $|v''b| = n$, it is also a factor of w', and no other letter x is such that $v''x \in \mathrm{Fact}(w')$. Hence $v = cv''b$ is the only right extension in w' of the factor $cv'' \neq Q$. $\square$

We can now proceed to prove Proposition 2.2.16.

Proof of Proposition 2.2.16. We first observe that u contains a single occurrence of Q. Indeed, if u contained other occurrences of Q, by Lemma 2.2.17 the suffix of u beginning with the penultimate occurrence would be a palindromic suffix of u strictly longer than Q, contradicting the hypothesis of maximality of the length of Q.

By Proposition 2.2.5 there exists an Arnoux-Rauzy word w such that $u \in \mathrm{Fact}(w)$. We can assume that $ua \notin \mathrm{Fact}(w)$ (otherwise ua is in $\mathrm{Fact}(AR)$ as required); so there exist $b \in A$ such that $b \neq a$ and $ub \in \mathrm{Fact}(w)$. Thus $aQb \in \mathrm{Fact}(w)$; since Q is a palindrome and $w \in AR$, also $bQa \in \mathrm{Fact}(w)$ and Q is a bispecial factor of w. Then it follows that every left special factor of w longer that Q must contain Q as a prefix, and since there is only a single occurrence of Q in u, Q itself is the longest suffix of u which is left special in w. Thus every occurrence of aQ in w must be "preceded" by s, i. e., if $w = \lambda aQ\mu$, then $w = \lambda' saQ\mu$, with $\lambda = \lambda' s$. In particular aQa is not a factor of w, for otherwise ua would be in $\mathrm{Fact}(w)$, contradicting our assumption.

Set $\Delta(w) = t_1 t_2 \cdots$. Let $B_1 = \varepsilon, B_2, \ldots$ be the sequence of all bispecial factors of w, ordered by increasing length, i. e., $|B_i| < |B_{i+1}|$ for all $i > 0$. By Proposition 2.2.4, they are the palindromic prefixes of w as w is not periodic. Moreover, for each $i > 0$ we have $B_{i+1} = (B_i t_i)^{(+)}$, so that $B_i t_i$ is left special and $t_i B_i$ is right special.

Since Q is a bispecial factor of w, one has $Q = B_m$ for some $m > 1$. Let $|Q| = n - 1$ for $n \geq 2$. We then have that $t_m Q$ is right special in w and, from Lemma 2.2.18, $t_m Qx \in \mathrm{Fact}(w)$ for all $x \in A$. It is clear that $t_m \neq a$ since $aQa \notin \mathrm{Fact}(w)$ and $t_m Qa \in \mathrm{Fact}(w)$, then we have that aQb and $t_m Qb$ are distinct factors of w, thus Qb is left special and bQ is the unique right special factor of w of length n. So $t_m = b$.

Let w' be any Arnoux-Rauzy sequence over A, whose directive word $\Delta(w') = t'_1 t'_2 \cdots$ satisfies $t'_i = t_i$ for $0 < i \leq m-1$ and $t'_m = a$. Since Q is the unique right special factor of w and w' of length $n-1$, from Lemma 2.2.19, we obtain that w and w' have the same factors of length k for each $k \leq n$. However, they differ on some factors of length $n+1$. Indeed, from the definition of w', we have that aQ is its unique right special factor of length n, so that by Lemma 2.2.18, for all $x \in A$ we have that $aQx \in \mathrm{Fact}(w')$. Therefore $aQa \in \mathrm{Fact}(w') \setminus \mathrm{Fact}(w)$.

Now let us prove that, as in w, each occurrence of aQ in w' is preceded by s. Let $p \in A^*$ be such that $|p| = |s|$ and $paQ \in \mathrm{Fact}(w')$. Let then S be the largest common suffix of paQ and saQ and Q' its prefix of length $n-1$. Clearly $Q \neq Q'$ since there is only one occurrence of Q in saQ. If we assume that $S \neq paQ$, then there exist $x, y \in A$ such that $x \neq y$, $xS \in \mathrm{Suff}(saQ)$ and $yS \in \mathrm{Suff}(paQ)$; then xQ' and yQ' are both factors of w and w' since these latter words have the same factors of length n. Thus Q' is a left special factor of w and w', and that is a contradiction, since the only left special factor of length $n-1$ in w and in w' is Q. Thus $p = s$ and so every occurrence of aQ in w' is preceded by s.

Since aQa is a factor of w', it follows that $saQa = ua$ is a factor of w'. Hence ua is in $\mathrm{Fact}(AR)$ as required. $\square$

From the preceding proposition one derives the following theorem, announced without proof in [27].

Theorem 2.2.20. *If w is a finite Episturmian word, then so is each of $w^{(+)}$ and $w^{(-)}$.*

Proof. Trivial if $w \in PAL$. Let then $w = a_1 \cdots a_n Q$, where $a_i \in A$ for $i = 1, \ldots, n$ and Q is the longest palindromic suffix of w. By Proposition 2.2.16, $wa_n = a_1 \cdots a_n Q a_n$ is a finite Episturmian word; since its longest palindromic suffix is $a_n Q a_n$, also $wa_n a_{n-1}$ is Episturmian. In this way, by applying Proposition 2.2.16 exactly n times, one eventually obtains that

$$a_1 a_2 \cdots a_n Q a_n \cdots a_2 a_1 = w^{(+)}$$

is Episturmian. Since $w^{(-)} = \tilde{w}^{(+)}$, the assertion follows. $\square$

Corollary 2.2.21. *Let $a \in A$ and $u \in A^*$. If au is a finite Episturmian word, then so is $au^{(+)}$.*

Proof. If au is not a palindrome, then by Theorem 2.2.20, $(au)^{(+)} = au^{(+)}a$ is an Episturmian word and therefore so is $au^{(+)}$. Let us then suppose that au is a palindrome.

By Theorem 2.2.20 one has $u^{(+)} \in \mathrm{Fact}(s)$ for a suitable $s \in AR$. Since s is recurrent there exist letters $x, y \in A$ such that

$$xu^{(+)}y \in \mathrm{Fact}(s) .$$

If $x \neq y$, then, since s is closed under reversal, one has also $yu^{(+)}x \in \mathrm{Fact}(s)$. Hence $u^{(+)}$ is bispecial, so that it follows $au^{(+)} \in \mathrm{Fact}(s)$. Let us now consider the case $x = y$. If $x = a$, then the assertion is trivially verified.

Suppose then $x \neq a$. As au is a palindrome, we can write $u = u'a$ with $u' \in PAL$. Hence,

$$x(u'a)^{(+)}x \in \mathrm{Fact}(s) .$$

Since $(u'a)^{(+)}$ begins with $u'a$ and ends with au', one has that $xu'a$ and $au'x$ are factors of s, so that u' is bispecial and then a palindromic prefix of s by Proposition 2.2.4.

Let $\Delta(s) = t_1 t_2 \cdots t_n \cdots$ be the directive word of s. There exists an integer k such that $u' = \psi(t_1 t_2 \cdots t_k)$. We consider any AR word s' whose directive word $\Delta(s')$ has the prefix $t_1 t_2 \cdots t_k a$. Thus $u'a = u$ is a prefix of s'. This implies, by Propositions 2.2.2 and 2.2.4, that $u^{(+)}$ is a bispecial prefix of s'. From this one derives $au^{(+)} \in \mathrm{Fact}(s')$. $\square$

2.3 Episturmian morphisms

We recall (cf. [29, 42, 43]) that a *standard Episturmian morphism* of A^* is any composition $\mu_w \circ \sigma$, with $w \in A^*$ and $\sigma : A^* \to A^*$ a morphism extending to A^* a permutation on the alphabet A. All these morphisms are injective. The set $\mathcal{E}$ of standard Episturmian morphisms is a monoid under map composition. The importance of standard Episturmian morphisms, and the reason for their name, lie in the following (see [29, 43]):

Theorem 2.3.1. *An injective morphism $\phi : A^* \to A^*$ is standard Episturmian if and only if $\phi(SEpi) \subseteq SEpi$, that is, if and only if it maps every standard Episturmian word over A into a standard Episturmian word over A.*

A *pure* standard Episturmian morphism is just a μ_w for some $w \in A^*$. Trivially, the set of pure standard Episturmian morphisms is the submonoid of $\mathcal{E}$ generated by the set $\{\mu_a \mid a \in A\}$. The following was proved in [29]:

Proposition 2.3.2. *Let $t \in A^\omega$ and $a \in A$. Then $\mu_a(t)$ is a standard Episturmian word if and only if so is t.*

Chapter 3

Episturmian words and generalizations

3.1 Involutory antimorphisms and pseudopalindromes

In this section we shall give the basic notions that, together with the Sturmian words, which have been presented before, will be the basis of the generalizations introduced in this chapter.

3.1.1 Antimorphisms of a free monoid

We recall that any (anti-)morphism whose domain is the free monoid A^* is uniquely determined by the images of the letters. Formally, for any monoid M and any map $\varphi : A \to M$, there exists a unique morphism $\hat{\varphi} : A^* \to M$ (resp. a unique antimorphism $\bar{\varphi} : A^* \to M$) that extends φ, i.e., such that $\hat{\varphi}|_A = \varphi$ (resp. $\bar{\varphi}|_A = \varphi$). This property characterizes free monoids, and is usually taken as the definition of free objects in the frame of category theory (cf. [47]).

A morphism or antimorphism $\varphi : A^* \to A^*$ is *involutory* if it is an involution of A^*, that is, if $\varphi^2 = \mathrm{id}$.

If $w = a_1 \cdots a_n \in A^*$, $a_i \in A$, $i = 1, \ldots, n$, the *mirror image*, or *reversal*, of w is the word

$$\tilde{w} = a_n \cdots a_1 \ .$$

One sets $\tilde{\varepsilon} = \varepsilon$. The map $R : A^* \to A^*$ defined by $w^R = \tilde{w}$ for any $w \in A^*$, called *reversal operator*, is clearly an involutory antimorphism of A^*.

Let τ be an involution of the alphabet A. Clearly, it can be regarded as a map $\tau : A \to A^*$, and then extended to a unique automorphism $\hat{\tau}$ of the free monoid A^*. The map $\vartheta = \hat{\tau} \circ R = R \circ \hat{\tau}$ is the unique involutory antimorphism of A^* extending the involution τ. One has, for $w = a_1 \cdots a_n$, $a_i \in A$, $i = 1, \ldots, n$,

$$w^\vartheta = a_n^\tau \cdots a_1^\tau \ .$$

Any involutory antimorphism of A^* can be constructed in this way; for example, the reversal R is obtained by extending the identity map of A.

If $A = \{a, b\}$, then there exist only two involutory antimorphisms, namely, the reversal R and the antimorphism $e = E \circ R$, called *exchange antimorphism*, extending the exchange map E defined on A as $E(a) = b$ and $E(b) = a$.

If the alphabet A has cardinality n, then the number of all involutory antimorphisms of A^* equals the number of the involutory permutations over n elements. As is well known, this number is given by

$$n! \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{1}{2^k (n-2k)! k!}$$

(sequence A000085 in [55]).

3.1.2 Pseudopalindromes

Let ϑ be an involutory antimorphism of A^*. A word $w \in A^*$ is called *ϑ-palindrome* if it is a fixpoint of ϑ, i.e., $w = w^\vartheta$. The set of all ϑ-palindromes of A^* is denoted by $PAL_\vartheta(A)$ or simply PAL_ϑ when there is no ambiguity.

An R-palindrome is usually called *palindrome* and PAL_R is denoted by PAL. In less precise terms, a word which is a ϑ-palindrome with respect to a given but unspecified involutory antimorphism ϑ, is also called *pseudopalindrome*.

Examples 3.1.1. The English word *racecar* is a palindrome.

Let $A = \{a, b\}$, e be the exchange antimorphism, and $w = abaabb$. One has $w^e = aabbab$. The word *abbaab* is an e-palindrome.

Let $A = \{a, b, c\}$ and τ be the involutory permutation defined as $\tau(a) = b$, $\tau(b) = a$, and $\tau(c) = c$. Setting $\vartheta = \tau \circ R$, the word *abcacbcab* is a ϑ-palindrome.

A word is called *ϑ-symmetric* if it is the product of two ϑ-palindromes. An R-symmetric word is simply called *symmetric*. In particular, any ϑ-palindrome is ϑ-symmetric.

Some combinatorial properties of symmetric words were studied in [23], and more recently in [10], where the term symmetric was used. One easily verifies that all words on the alphabet $\{a, b\}$ of length ≤ 5 are symmetric. The word $w = abaabb$ is not symmetric but it is e-symmetric, because it is the product of the two words ab and $aabb$ which are e-palindromes.

In the remaining part of this section, we will assume that ϑ is a fixed involutory antimorphism of A^*. To simplify the notation, for any $w \in A^*$, we shall denote by $\bar{w}$ the word w^ϑ, so that for all $u, v \in A^*$ one has

$$|\bar{u}| = |u|\,, \quad \overline{uv} = \bar{v}\bar{u}\,, \quad \text{and} \quad \bar{\bar{u}} = u\,.$$

Lemma 3.1.2. *A word w is a conjugate of $\bar{w}$ if and only if it is ϑ-symmetric.*

Proof. If $w = \alpha\beta$ with $\alpha, \beta \in PAL_\vartheta$, then $\bar{w} = \beta\alpha$, so that $w \sim \bar{w}$. Conversely, suppose that w and $\bar{w}$ are conjugate. One can write $w = \lambda\mu$ and $\bar{w} = \mu\lambda$ for some $\lambda, \mu \in A^*$. Thus $w = \bar{\lambda}\bar{\mu} = \lambda\mu$. Since $|\lambda| = |\bar{\lambda}|$, one obtains $\lambda = \bar{\lambda}$ and $\mu = \bar{\mu}$. □

Lemma 3.1.3. *A ϑ-palindrome $w \in A^+$ has a period $p \leq |w|$ if and only if it has a ϑ-palindromic prefix (suffix) of length $|w| - p$.*

Proof. If w has a period $p \leq |w|$, then it has a border v of length $|w| - p$, so that we can write $w = \lambda v = v\mu$ for some words λ and μ. Since w is a ϑ-palindrome, one has

$$w = v\mu = \bar{v}\bar{\lambda}\,.$$

Therefore, $v = \bar{v}$. Conversely, if the ϑ-palindrome w has the ϑ-palindromic prefix v, one has

$$w = v\mu = \bar{\mu}v\,,$$

so that v is a border of w and $|w| - |v|$ is a period of w. □

Lemma 3.1.4. *Let $w \in A^+$ and z_w be its fractional root. The word $z_{\bar{w}}$ is a conjugate of $\bar{z}_w$.*

Proof. Let w be a nonempty word. Since ϑ acts on the alphabet as a permutation, one derives that p is a period of w if and only if it is a period of $\bar{w}$. Therefore one has $\pi_w = \pi_{\bar{w}}$. We can write $w = z_w^k z'$ with $k \geq 1$ and z' a proper prefix of z_w, and

$$\bar{w} = \bar{z}'\bar{z}_w^k = z_{\bar{w}}^h\, z''$$

with $h \geq 1$ and z'' a proper prefix of $z_{\bar{w}}$. Since $|w| = |\bar{w}|$ and $|\bar{z}_w| = |z_{\bar{w}}| = \pi_{\bar{w}}$, one has $h = k$ and, by Lemma 2.1.32, $\bar{z}_w \sim z_{\bar{w}}$. □

The following lemma, which will be useful in the sequel, can be derived by the propositions above.

Lemma 3.1.5. *Let $w \in A^+$ be a ϑ-palindrome having a period $p \leq |w|$. Any factor u of w of length p is ϑ-symmetric. In particular, z_w is ϑ-symmetric.*

Proof. Since $w = \bar{w}$ and $|u| = p$, by Lemma 2.1.32 one has $u \sim \bar{u}$. Hence, by Lemma 3.1.2 one obtains $u \in PAL_\vartheta^2$. As $|z_w| = \pi_w$, one derives $z_w \in PAL_\vartheta^2$. □

For instance, let $A = \{a, b\}$ and let $\vartheta(a) = b$, $\vartheta(b) = a$. The word $w = babaababbaba$ is a ϑ-palindrome, having the periods 8 and 10. Any factor of w of length 8 or 10 belongs to PAL_ϑ^2; as an example, $abaababb = (ab)(aababb) \in PAL_\vartheta^2$.

For any involutory antimorphism ϑ, one can define the (right) ϑ-palindrome closure operator: for any $w \in A^*$, $w^{\oplus_\vartheta}$ denotes the shortest ϑ-palindrome having w as a prefix.

We shall drop the subscript ϑ from the ϑ-palindrome closure operator ${}^{\oplus_\vartheta}$ when no confusion arises. As one easily verifies (cf. [27]), if Q is the longest ϑ-palindromic suffix of w and $w = sQ$, then

$$w^{\oplus} = sQ\bar{s}\;.$$

Example 3.1.6. Let $A = \{a, b, c\}$ and ϑ be defined as $\bar{a} = b$, $\bar{c} = c$. If $w = abacabc$, then $Q = cabc$ and $w^{\oplus} = abacabcbab$.

The following lemma will be useful in the sequel.

Lemma 3.1.7. *For any $u \in PAL_\vartheta \setminus \{\varepsilon\}$ and $a \in A$, $(ua)^{\oplus}$ is a first return to u, i. e. , if $(ua)^{\oplus} = \lambda u \rho$ with $\lambda, \rho \in A^*$, then either $\lambda = \varepsilon$ or $\rho = \varepsilon$.*

Proof. By contradiction, let $\lambda, \rho \in A^+$ be such that

$$(ua)^{\oplus} = \lambda u \rho \, . \tag{3.1}$$

Clearly $|\lambda| + |u| + |\rho| = |(ua)^{\oplus}| \leq 2|u| + 2$, which implies $|\lambda| \leq |u| + 2 - |\rho| \leq |u| + 1$. Let us show that actually one has $|\lambda| \leq |u|$. Indeed, if $\lambda = ua$ then from (3.1) one derives $|(ua)^{\oplus}| = 2|u| + 2$; this implies that $a \notin PAL_{\vartheta}$ and $(ua)^{\oplus} = ua\bar{a}u = uau\rho$, so that $u\rho = \bar{a}u$. It follows that for some $k > 0$, $u = \bar{a}^k \notin PAL_{\vartheta}$, a contradiction.

Let then $v, w \in A^*$ be such that $u = \lambda v$ and $(ua)^{\oplus} = uw = \bar{w}u$, whence $\lambda u \rho = uw = \lambda v w$. Thus $u\rho = vw$, so that v is also a prefix of u and therefore a border of u. Since u is a ϑ-palindrome, v is a ϑ-palindrome too, so that $u = \lambda v = v\bar{\lambda}$. Therefore

$$(ua)^{\oplus} = \lambda u \rho = \lambda v \bar{\lambda} \rho \, .$$

Thus $\lambda v \bar{\lambda}$ is a ϑ-palindrome beginning with ua and strictly shorter than $(ua)^{\oplus}$, which is a contradiction. ☐

3.1.3 Unbordered pseudopalindromes

We denote by $\mathcal{P}_{\vartheta}$ the set of unbordered ϑ-palindromes. We remark that $\mathcal{P}_{\vartheta}$ is a *biprefix code*. This means that every word of $\mathcal{P}_{\vartheta}$ is neither a prefix nor a suffix of any other element of $\mathcal{P}_{\vartheta}$. We observe that $\mathcal{P}_R = A$. The following result was proved in [12]:

Proposition 3.1.8 (Theorem 3.6.2). $PAL_{\vartheta}^* = \mathcal{P}_{\vartheta}^*$.

This can be equivalently stated as follows: every ϑ-palindrome can be uniquely factorized by the elements of $\mathcal{P}_{\vartheta}$. For instance, the ϑ-palindrome $abacabcbab$ of Example 3.1.6 is factorizable as $ab \cdot acabcb \cdot ab$, with $acabcb, ab \in \mathcal{P}_{\vartheta}$.

Since $\mathcal{P}_{\vartheta}$ is a code, the map

$$\begin{array}{rcl} f : \mathcal{P}_{\vartheta} & \longrightarrow & A \\ \pi & \longmapsto & \pi^f \end{array} \tag{3.2}$$

can be extended (uniquely) to a morphism $f : \mathcal{P}_{\vartheta}^* \to A^*$. Moreover, since $\mathcal{P}_{\vartheta}$ is a prefix code, any word in $\mathcal{P}_{\vartheta}^{\omega}$ can be uniquely factorized by the elements of $\mathcal{P}_{\vartheta}$, so that f can be naturally extended to $\mathcal{P}_{\vartheta}^{\omega}$.

Proposition 3.1.9. *Let $\phi : X^* \to A^*$ be an injective morphism such that $\phi(X) \subseteq \mathcal{P}_\vartheta$. Then, for any $w \in X^*$:*

1. $\phi(\tilde{w}) = \overline{\phi(w)}$,
2. $w \in PAL \iff \phi(w) \in PAL_\vartheta$,
3. $\phi(w^{(+)}) = \phi(w)^\oplus$.

Proof. The first statement is trivially true for $w = \varepsilon$. If $w = x_1 \cdots x_n$ with $x_i \in X$ for $i = 1, \ldots, n$, then since $\phi(X) \subseteq \mathcal{P}_\vartheta \subseteq PAL_\vartheta$,

$$\phi(\tilde{w}) = \phi(x_n) \cdots \phi(x_1) = \overline{\phi(x_n)} \cdots \overline{\phi(x_1)} = \overline{\phi(w)} \ .$$

As ϕ is injective, statement *2* easily follows from *1*.

Finally, let $\phi(w) = vQ$ where $v \in A^*$ and Q is the longest ϑ-palindromic suffix of $\phi(w)$. Since $\phi(w), Q \in \mathcal{P}_\vartheta^*$ and $\mathcal{P}_\vartheta$ is a biprefix code, we have $v \in \mathcal{P}_\vartheta^*$. This implies, as ϕ is injective, that there exist $w_1, w_2 \in X^*$ such that $w = w_1 w_2$, $\phi(w_1) = v$, and $\phi(w_2) = Q$. By *2*, w_2 is the longest palindromic suffix of w. Hence, by *1*:

$$\phi(w^{(+)}) = \phi(w_1 w_2 \tilde{w}_1) = vQ\bar{v} = \phi(w)^\oplus \ ,$$

as desired. □

Example 3.1.10. Let $X = \{a, b, c\}$, $A = \{a, b, c, d, e\}$, and ϑ be defined in A as $\bar{a} = b$, $\bar{c} = c$, and $\bar{d} = e$. Let $\phi : X^* \to A^*$ be the injective morphism defined by $\phi(a) = ab$, $\phi(b) = ba$, $\phi(c) = dce$. One has $\phi(X) \subseteq \mathcal{P}_\vartheta$ and

$$\phi\left((abc)^{(+)}\right) = \phi(abcba) = abbadcebaab = (\phi(abc))^\oplus \ .$$

3.2 Pseudostandard words

We can naturally define a map $\psi_\vartheta : A^* \to A^*$ by $\psi_\vartheta(\varepsilon) = \varepsilon$ and

$$\psi_\vartheta(ua) = (\psi_\vartheta(u)a)^\oplus$$

for $u \in A^*$, $a \in A$. For any $u, v \in A^*$ one has $\psi_\vartheta(uv) \in \psi_\vartheta(u)A^* \cap A^*\psi_\vartheta(u)$, so that, as done for the iterated palindrome closure, the domain of ψ_ϑ can be extended to infinite words too. More precisely, if $x \in A^\omega$, then

$$\psi_\vartheta(x) = \lim_{n\to\infty} \psi_\vartheta(w_n) \ ,$$

where $\{w_n\} = \mathrm{Pref}(x) \cap A^n$ for all $n \geq 0$. The word x is called the *directive word* of $\psi_\vartheta(x)$ and is denoted by $\Delta(\psi_\vartheta(x))$. The images of infinite words over A by ψ_ϑ have been called *ϑ-standard words* in [27]. If $\vartheta = R$, then $\psi_R = \psi$, where ψ is the iterated palindrome closure operator introduced before, so that an R-standard word is a standard Episturmian word. A ϑ-standard word, without specifying the antimorphism ϑ, has been called *pseudostandard word.*

Example 3.2.1. Let $A = \{a, b\}$ and $\vartheta = E \circ R$, so that $\bar{a} = b$. For $x = (ab)^\omega$, we have $\psi_\vartheta(a) = ab$, $\psi_\vartheta(ab) = abbaab$, and

$$s = \psi_\vartheta(x) = abbaababbaabbaab \cdots .$$

The word s is the ϑ-standard word having x as its directive word.

The following theorem, proven in [27], shows that any ϑ-standard word is a morphic image of the standard Episturmian word having the same directive word.

Theorem 3.2.2. *For any $w \in A^\infty$, one has $\psi_\vartheta(w) = \mu_\vartheta(\psi(w))$, where μ_ϑ is the injective morphism defined as $\mu_\vartheta(a) = a^\oplus$ for any letter $a \in A$.*

For instance, one easily verifies that the word s of Example 3.2.1 is equal to $\mu(f)$, where f is the Fibonacci word and $\mu = \mu_\vartheta$ is the Thue-Morse morphism defined as $\mu(a) = ab$, $\mu(b) = ba$.

A new proof of Theorem 3.2.2 will be given in Section 3.3, as a consequence of a more general result. Some general properties of ϑ-standard words have been considered in [27]. In particular, we recall that

Proposition 3.2.3. *Let $s = \psi_\vartheta(x)$ be a ϑ-standard word. The following hold:*

1. *w is a prefix of s if and only if $w^\oplus$ is a prefix of s,*

2. *the set of all ϑ-palindromic prefixes of s is given by $\psi_\vartheta(\mathrm{Pref}(x))$,*

3. *s is closed under ϑ, i. e. , if $w \in \mathrm{Fact}(s)$, then $\bar{w} \in \mathrm{Fact}(s)$.*

Moreover, the following holds:

Proposition 3.2.4. *If s is a ϑ-standard word over A and two letters of A occur infinitely often in $\Delta(s)$, then any prefix of s is a left special factor of s.*

Proof. A prefix p of s is also a prefix of any ϑ-palindromic prefix B of s such that $|p| \leq |B|$. Since B is a suffix of any ϑ-palindromic prefix of s whose length is at least $|B|$, and there exist two distinct letters (say a and b) which occur infinitely often in $\Delta(s)$, by Proposition 3.2.3 one derives $Ba, Bb \in \mathrm{Fact}(s)$. Therefore, as $\bar{p} \in \mathrm{Suff}(B)$, we have $\bar{p}a, \bar{p}b \in \mathrm{Fact}(s)$, i. e. , $\bar{p}$ is right special. Since by Proposition 3.2.3 s is closed under ϑ, one has $\bar{a}p, \bar{b}p \in \mathrm{Fact}(s)$; as $\bar{a} \neq \bar{b}$, p is left special. □

For the converse of the previous proposition, we observe that a ϑ-standard word s can have left special factors which are not prefixes of s. For instance, consider the ϑ-standard word s in Example 3.2.1. As one easily verifies, b and ba are two left special factors of s, which are not prefixes.

However, we will show that if a left special factor w of a ϑ-standard word s is not a prefix of s, then $|w| \leq 2$. For a proof of this we need a couple of lemmas. We denote by $A' = A \setminus PAL_\vartheta$ the set of letters of A that are not ϑ-palindromic.

Lemma 3.2.5. *The following holds:*

$$A'\mu_\vartheta(A^*) \cap \mu_\vartheta(A^*) = \mu_\vartheta(A^*)A' \cap \mu_\vartheta(A^*) = \emptyset\ .$$

Proof. It is sufficient to observe that any word in $\mu_\vartheta(A^*)$ has an even number of occurrences of letters in A'. □

Lemma 3.2.6. *Let $b, c \in A'$, and let $f = \bar{b}\mu_\vartheta(u)$ and $g = \mu_\vartheta(v)c$ be factors of a ϑ-standard word $t = \mu_\vartheta(s)$, with $s \in SEp$. Then:*

1. *If $bu, vc \in \mathrm{Fact}(s)$ and $|f| > 1$, then $f \neq g$.*

2. *If $u \in \mathrm{Fact}(s)$ and $|f| > 3$, then $bu \in \mathrm{Fact}(s)$.*

Proof. (1). Since $|f| > 1$, one has $u \neq \varepsilon$. By contradiction, if $f = g$, one has also $v \neq \varepsilon$, so that, from the definition of μ_ϑ, $\bar{b}b$ is a prefix of $\mu_\vartheta(v)$. Then $b\bar{b}$ is a prefix of $\mu_\vartheta(u)$, and so on; therefore, $f = \bar{b}(b\bar{b})^k = (\bar{b}b)^k\bar{b}$ for $k = |u| = |v| \geq 1$. Hence $c = \bar{b}$, $u = b^k$, and $v = \bar{b}^k$. As $k \geq 1$, by Proposition 2.2.6, $bu = b^{k+1}$ and $vc = \bar{b}^{k+1}$ cannot be both factors of the Episturmian word s, a contradiction.

(2). Since $|f| > 3$, one derives $|u| > 1$. By contradiction, suppose $bu \notin \mathrm{Fact}(s)$. By the preceding lemma and by Theorem 3.2.2, one derives $f =$

$\mu_\vartheta(v')c'$ for some suitable $v' \in A^*$ and $c' \in A'$ such that $v'c' \in \text{Fact}(s)$. As done before, one then obtains $f = (\bar{b}b)^k\bar{b}$ so that $b^k, \bar{b}^k \in \text{Fact}(s)$, which is absurd by Proposition 2.2.6, as $k \geq 2$. $\square$

Theorem 3.2.7. *Let w be a left special factor of a ϑ-standard word $t = \mu_\vartheta(s)$, with $s \in SEp$. If $|w| \geq 3$, then w is a prefix of t.*

Proof. By Theorem 3.2.2, w can be written in one of the following ways:

1. $w = \mu_\vartheta(u)$, with $u \in \text{Fact}(s)$,
2. $w = \bar{b}\mu_\vartheta(u)$, with $bu \in \text{Fact}(s)$ and $b \in A'$,
3. $w = \mu_\vartheta(u)c$, with $uc \in \text{Fact}(s)$ and $c \in A'$,
4. $w = \bar{b}\mu_\vartheta(u)c$, with $buc \in \text{Fact}(s)$ and $b, c \in A'$.

In case 1, let $xw, yw \in \text{Fact}(t)$ with $x \neq y$ letters of A. If x is ϑ-palindromic, then clearly one must have $xu \in \text{Fact}(s)$. If $x \in A'$, then by the preceding lemma one has $\bar{x}u \in \text{Fact}(s)$, as $|xw| > 3$. Since the same holds for y, u is a left special factor of the Episturmian word s, and therefore a prefix of it. Thus $w = \mu_\vartheta(u)$ is a prefix of t.

Cases 2 and 4 are absurd; indeed, by the preceding lemma one derives that every occurrence of w is preceded by b.

Finally, in case 3, by the preceding lemma one derives that every occurrence of w is followed by $\bar{c}$. Hence $\mu_\vartheta(uc)$ is a left special factor of t and one can apply the same argument as in case 1 to show that it is a prefix of t. $\square$

An infinite word t is a *ϑ-word* if there exists a ϑ-standard word s such that $\text{Fact}(t) = \text{Fact}(s)$. An R-word is an Episturmian word.

Proposition 2.2.16 and Theorem 2.2.20 can be extended to the class of ϑ-words, showing that if w is a factor of a ϑ-word, then $w^\oplus$ and $w^\ominus$ are also factors of ϑ-words. A proof can be obtained as a consequence of Theorems 2.2.20 and 3.2.2 and of Corollary 2.2.21. However, we need the following lemma (cf. [27]):

Lemma 3.2.8. *Let $u \in A^*$ and $x \in A \cup \{\varepsilon\}$. Then*

$$(\mu_\vartheta(u)x)^\oplus = \mu_\vartheta\left((ux)^{(+)}\right) .$$

Theorem 3.2.9. *Let w be a factor of a ϑ-standard word. Then each of $w^{\oplus}$ and $w^{\ominus}$ is a factor of a ϑ-standard word.*

Proof. We shall suppose $w \notin PAL_{\vartheta}$, otherwise the result is trivial. Since $w^{\ominus} = \bar{w}^{\oplus}$, it suffices to prove the result for $w^{\oplus}$. Let $A' = A \setminus PAL_{\vartheta}$ as above. From Theorem 3.2.2, one derives that w can be written in one of the following ways:

1. $w = \mu_{\vartheta}(u)x$, with $x \in A \cup \{\varepsilon\}$ and $ux \in \text{Fact}(Ep)$,
2. $w = \bar{a}\mu_{\vartheta}(u)b$, with $a, b \in A'$ and $aub \in \text{Fact}(Ep)$,
3. $w = \bar{a}\mu_{\vartheta}(u)$, with $a \in A'$ and $au \in \text{Fact}(Ep)$.

In the first case, by Theorem 2.2.20 there exists a standard Episturmian word $s = \psi(\Delta)$ such that $(ux)^{(+)} \in \text{Fact}(s)$. Thus, by Lemma 3.2.8 and Theorem 3.2.2, $w^{\oplus} = \mu_{\vartheta}\left((ux)^{(+)}\right)$ is a factor of the ϑ-standard word $\psi_{\vartheta}(\Delta) = \mu_{\vartheta}(s)$.

In the second case, by using Lemma 3.2.8, one has:

$$w^{\oplus} = \bar{a}\left(\mu_{\vartheta}(u)b\right)^{\oplus} a = \bar{a}\mu_{\vartheta}\left((ub)^{(+)}\right) a \in \text{Fact}\left(\mu_{\vartheta}\left(a(ub)^{(+)}a\right)\right) .$$

Moreover, aub is not a palindrome, since otherwise one would derive, for instance using Lemma 3.2.8, that $w = \bar{a}\mu_{\vartheta}(u)b$ is a ϑ-palindrome, which contradicts our assumption. Thus $(aub)^{(+)} = a(ub)^{(+)}a$ and the result is a consequence of Theorem 3.2.2.

In the third case, since w is not a ϑ-palindrome, by Lemma 3.2.8 one obtains

$$w^{\oplus} = \bar{a}\mu_{\vartheta}(u)^{\oplus}a \in \text{Fact}\left(\mu_{\vartheta}(au^{(+)}a)\right) .$$

If $u = a^k$ for some $k \geq 0$, then $au^{(+)}a = a^{k+2} \in \text{Fact}(Ep)$; otherwise $au^{(+)}$ is not a palindrome and $au^{(+)}a = (au^{(+)})^{(+)}$, so that $au^{(+)}a$ is Episturmian by Corollary 2.2.21 and Theorem 2.2.20. Once again, the assertion follows from Theorem 3.2.2. □

Corollary 3.2.10. *Let w be a factor of a ϑ-standard word. Then there exists a ϑ-standard word having both $w^{\oplus}$ and $w^{\ominus}$ as factors.*

Proof. Trivial if $w \in PAL_\vartheta$. Let then $w = Pbt = saQ$, where P (resp. Q) is the longest ϑ-palindromic prefix (resp. suffix) of w, and $a, b \in A$. Thus $w\bar{a}$ and $\bar{b}w$, being respectively factors of $w^\oplus = saQ\bar{a}\bar{s}$ and $w^\ominus = \bar{t}\bar{b}Pbt$, are factors of ϑ-standard words by Theorem 3.2.9.

Suppose $w\bar{a} \notin PAL_\vartheta$. Then $(w\bar{a})^\ominus = aw^\ominus\bar{a}$, so that $w^\ominus\bar{a}$ is a factor of some ϑ-standard word, by Theorem 3.2.9. Consider the word

$$(w^\ominus\bar{a})^\oplus = (\bar{t}\bar{b}Pbt\bar{a})^\oplus = (\bar{t}\bar{b}saQ\bar{a})^\oplus ,$$

and call Q' the longest ϑ-palindromic suffix of $w^\ominus\bar{a}$; then $Q' = aQ\bar{a}$. Indeed, since $aQ\bar{a}$ is a ϑ-palindrome, one has $|Q'| \geq |aQ\bar{a}|$; but $|aQ\bar{a}| < |Q'| \leq |saQ\bar{a}|$ is absurd, for Q would not be the longest ϑ-palindromic suffix of w, and $|Q'| > |saQ\bar{a}|$ cannot happen, for otherwise there would exist a ϑ-palindromic proper suffix of $w^\ominus$ having w as a suffix, contradicting the definition of $w^\ominus$. Thus

$$(w^\ominus\bar{a})^\oplus = \bar{t}\bar{b}saQ\bar{a}\bar{s}bt = \bar{t}\bar{b}Pbt\bar{a}\bar{s}bt$$

is a factor of some ϑ-standard word, again by Theorem 3.2.9, and it contains both $w^\oplus$ and $w^\ominus$ as factors.

If $w\bar{a} \in PAL_\vartheta$ but $\bar{b}w \notin PAL_\vartheta$, one can prove by a symmetric argument that $(\bar{b}w^\oplus)^\ominus$ is a factor of some ϑ-standard word having both $w^\oplus$ and $w^\ominus$ as factors. Let then $w\bar{a}, \bar{b}w \in PAL_\vartheta$, so that

$$w^\oplus = w\bar{a} = a\bar{w} \;\text{ and } w^\ominus = \bar{b}w = \bar{w}b . \tag{3.3}$$

If w is a single letter, one derives $w = a = b$, so that $w^\oplus = a\bar{a}$ and $w^\ominus = \bar{a}a$. Therefore $w^\oplus$ and $w^\ominus$ are factors of any ϑ-standard word whose directive word begins with a^2. Let us then suppose $|w| > 1$. From (3.3) it follows $w = aRb$ for some $R \in A^*$ such that $aR = \bar{R}\bar{a} = P$ and $Rb = \bar{b}\bar{R} = Q$. Moreover,

$$w = aRb = a\bar{b}\bar{R} = \bar{R}\bar{a}b , \tag{3.4}$$

showing that $\bar{R}$ is a border of w. Therefore one has either $w = (a\bar{b})^k$ or $w = (a\bar{b})^k a$, for some $k > 0$. In the first case, from (3.4) one derives $a = \bar{a}$ and $b = \bar{b}$, so that any ϑ-standard word whose directive word begins with ab^{k+1} contains both $w^\oplus = (ab)^k a$ and $w^\ominus = b(ab)^k$ as factors. In the latter case, by (3.4) one obtains $a = b$, so that any ϑ-standard word whose directive word begins with a^{k+1} contains both $w^\oplus = (a\bar{a})^k$ and $w^\ominus = (\bar{a}a)^k$ as factors. $\square$

Remark. For a finite Episturmian word w, the proof of the preceding result can be simplified by using Theorem 2.2.20 and Corollary 2.2.21. Indeed, if w is not a palindrome, we can write $w = Pbt = saQ$, where P and Q are respectively the longest palindromic prefix and suffix of w, and $a, b \in A$. By Theorem 2.2.20, $w^{(+)}$ and $w^{(-)}$ are finite Episturmian words; moreover bw is a factor of $w^{(-)}$, so that by Corollary 2.2.21, $bw^{(+)}$ is a finite Episturmian word. By Theorem 2.2.20, $\left(bw^{(+)}\right)^{(-)}$ is a finite Episturmian word, which has also $w^{(-)}$ as a factor, as one can prove similarly as in the proof of Corollary 3.2.10.

In the case of Sturmian words, results analogous to Theorem 3.2.9 and Corollary 3.2.10 were proven in [27] with a different and simpler technique based on the structure of finite Sturmian words.

Example 3.2.11. Let τ be the Tribonacci word

$$\tau = \psi\left((abc)^\omega\right) = abacabaabacababacabaabacabac\cdots .$$

If $w = bac \in \mathrm{Fact}(\tau)$, one has that $w^{(+)} = bacab$ and $w^{(-)} = cabac$ are factors of τ. However, in the case of the factor $v = abacabab$, one has $v^{(+)} = abacababacaba \in \mathrm{Fact}(\tau)$, whereas $v^{(-)} = babacabab$ is not a factor of τ, since otherwise v would be a left special factor of τ, which is a contradiction as $v \notin \mathrm{Pref}(\tau)$. Nevertheless, both $v^{(+)}$ and $v^{(-)}$ are factors of any Episturmian word whose directive word begins with $abcbb$. Indeed, $v = Pb$ where $P = abacaba$ is the longest palindromic prefix of v, and

$$\left(bv^{(+)}\right)^{(-)} = abacababacababacaba = \psi(abcbb) .$$

3.3 Words generated by nonempty seeds

We now consider a generalization of the construction of ϑ-standard words. Define the map $\hat{\psi}_\vartheta : A^* \to A^*$ by setting $\hat{\psi}_\vartheta(\varepsilon) = u_0$ with u_0 a fixed word of A^* called *seed*, and

$$\hat{\psi}_\vartheta(ua) = \left(\hat{\psi}_\vartheta(u)a\right)^\oplus$$

for $u \in A^*$ and $a \in A$. As usual, we can extend this definition to infinite words $t \in A^\omega$ by:

$$\hat{\psi}_\vartheta(t) = \lim_{n\to\infty} \hat{\psi}_\vartheta(w_n) ,$$

where $\{w_n\} = \mathrm{Pref}(t) \cap A^n$ for all $n \geq 0$. The word t is called the *directive word* of $\hat{\psi}_\vartheta(t)$, and denoted by $\Delta(\hat{\psi}_\vartheta(t))$. When the seed u_0 is empty, one has $\hat{\psi}_\vartheta = \psi_\vartheta$ so that one obtains ϑ-standard words. If $u_0 \neq \varepsilon$, then any word $\hat{\psi}_\vartheta(t)$ is called *ϑ-standard with seed.*

Example 3.3.1. Let $A = \{a, b, c\}$, ϑ be the involutory antimorphism exchanging a and b and fixing c, $u_0 = acbbc$, and $w = abc$. Then

$$\begin{aligned}\hat{\psi}_\vartheta(w) &= \left(\hat{\psi}_\vartheta(ab)c\right)^\oplus = \left(\left(\hat{\psi}_\vartheta(a)b\right)^\oplus c\right)^\oplus = \left(\left((acbbca)^\oplus b\right)^\oplus c\right)^\oplus \\ &= \left((acbbcaacbb)^\oplus c\right)^\oplus = acbbcaacbbcaacbcacbbcaacbbcaacb\ .\end{aligned}$$

Example 3.3.2. Let $A = \{a, b, c\}$, ϑ be the involutory antimorphism exchanging b and c and fixing a, $u_0 = ab$, and $w = aac$. Then

$$\begin{aligned}\hat{\psi}_\vartheta(w) &= \left(\hat{\psi}_\vartheta(aa)c\right)^\oplus = \left(\left(\hat{\psi}_\vartheta(a)a\right)^\oplus c\right)^\oplus = \left((abacaa)^\oplus c\right)^\oplus \\ &= (abacaabacac)^\oplus = abacaabacacbabacaabaca\ .\end{aligned}$$

Let $t = xt_1t_2\cdots$, with $x \in A$ and $t_i \in A$ for $i \geq 1$. We remark that the set of ϑ-palindromic prefixes of the word $w = \hat{\psi}_\vartheta(t)$ is

$$(PAL_\vartheta \cap \mathrm{Pref}(u_0)) \cup \{u_n \mid n \geq 1\}\ ,$$

where $u_1 = (u_0x)^\oplus$ and $u_{i+1} = (u_it_i)^\oplus$ for $i \geq 1$.

Define the endomorphism ϕ_x of A^* by setting

$$\phi_x(a) = \hat{\psi}_\vartheta(xa)\hat{\psi}_\vartheta(x)^{-1}$$

for any letter $a \in A$. From the definition, one has that ϕ_x depends on ϑ and u_0; moreover, $\phi_x(a)$ ends with $\bar{a}$ for all $a \in A$, so that any word of the set $X = \phi_x(A)$ is uniquely determined by its last letter. Thus X is a suffix code, and ϕ_x is an injective morphism.

Example 3.3.3. Let A, ϑ, and u_0 be defined as in Example 3.3.1, and let $x = a$. Then

$$\begin{aligned}\phi_a(a) &= \hat{\psi}_\vartheta(aa)\hat{\psi}_\vartheta(a)^{-1} = acbbcaacb\ , \\ \phi_a(b) &= \hat{\psi}_\vartheta(ab)\hat{\psi}_\vartheta(a)^{-1} = acbbca\ , \\ \phi_a(c) &= \hat{\psi}_\vartheta(ac)\hat{\psi}_\vartheta(a)^{-1} = acbbcaacbc\ .\end{aligned} \tag{3.5}$$

To simplify the notation, in the following we shall often omit in the proofs the subscript x from ϕ_x, when no confusion arises.

Theorem 3.3.4. *Fix $x \in A$ and $u_0 \in A^*$. Let $\hat{\psi}_\vartheta$ and ϕ_x be defined as above. Then for any $w \in A^*$, the following holds:*

$$\hat{\psi}_\vartheta(xw) = \phi_x(\psi(w))\hat{\psi}_\vartheta(x) \ .$$

Proof. In the following we shall often use the property that if γ is an endomorphism of A^* and v is a suffix of $u \in A^*$, then $\gamma(uv^{-1}) = \gamma(u)\gamma(v)^{-1}$.

We will prove the theorem by induction on $|w|$. It is trivial that for $w = \varepsilon$ the claim is true since $\psi(\varepsilon) = \varepsilon = \phi(\varepsilon)$. Suppose that for all the words shorter than w, the statement holds. For $|w| > 0$, we set $w = vy$ with $y \in A$.

First we consider the case $|v|_y \neq 0$. We can then write $v = v_1yv_2$ with $|v_2|_y = 0$, so that

$$\hat{\psi}_\vartheta(xv) = \hat{\psi}_\vartheta(xv_1yv_2) = \hat{\psi}_\vartheta(xv_1)y\lambda = \bar{\lambda}\bar{y}\hat{\psi}_\vartheta(xv_1) \ ,$$

for a suitable $\lambda \in A^*$. Note that $\hat{\psi}_\vartheta(xv_1)$ is the largest ϑ-palindromic prefix (resp. suffix) followed (resp. preceded) by y (resp. $\bar{y}$) in $\hat{\psi}_\vartheta(xv)$. Therefore,

$$\hat{\psi}_\vartheta(xvy) = \bar{\lambda}\bar{y}\hat{\psi}_\vartheta(xv_1)y\lambda = \hat{\psi}_\vartheta(xv)\hat{\psi}_\vartheta(xv_1)^{-1}\hat{\psi}_\vartheta(xv) \ . \tag{3.6}$$

By a similar argument one has:

$$\psi(vy) = \psi(v)\psi(v_1)^{-1}\psi(v) \ . \tag{3.7}$$

By induction we have:

$$\hat{\psi}_\vartheta(xv) = \phi(\psi(v))\hat{\psi}_\vartheta(x) \ , \quad \hat{\psi}_\vartheta(xv_1) = \phi(\psi(v_1))\hat{\psi}_\vartheta(x) \ .$$

Replacing in (3.6), and by (3.7), we obtain

$$\begin{aligned} \hat{\psi}_\vartheta(xvy) &= \phi(\psi(v))\phi(\psi(v_1))^{-1}\phi(\psi(v))\hat{\psi}_\vartheta(x) \\ &= \phi(\psi(v)\psi(v_1)^{-1}\psi(v))\hat{\psi}_\vartheta(x) \\ &= \phi(\psi(vy))\hat{\psi}_\vartheta(x) \ , \end{aligned}$$

which was our aim.

Now suppose that $|v|_y = 0$ and $PAL_\vartheta \cap \mathrm{Pref}(u_0x)y^{-1} \neq \emptyset$. Let α_y be the longest word in $PAL_\vartheta \cap \mathrm{Pref}(u_0x)y^{-1}$, that is the longest ϑ-palindromic prefix of u_0x which is followed by y. Since $|v|_y = 0$, one derives that the longest ϑ-palindromic suffix of $\hat{\psi}_\vartheta(xv)y$ is $\bar{y}\alpha_y y$, whence

$$\hat{\psi}_\vartheta(xvy) = \left(\hat{\psi}_\vartheta(xv)y\right)^{\oplus} = \hat{\psi}_\vartheta(xv)\alpha_y^{-1}\hat{\psi}_\vartheta(xv) \ . \tag{3.8}$$

By induction, this implies

$$\hat{\psi}_\vartheta(xvy) = \phi(\psi(v))\hat{\psi}_\vartheta(x)\alpha_y^{-1}\phi(\psi(v))\hat{\psi}_\vartheta(x) \ . \tag{3.9}$$

By using (3.8) for $v = \varepsilon$, one has $\hat{\psi}_\vartheta(xy) = \hat{\psi}_\vartheta(x)\alpha_y^{-1}\hat{\psi}_\vartheta(x)$, and

$$\phi(y) = \hat{\psi}_\vartheta(xy)\left(\hat{\psi}_\vartheta(x)\right)^{-1} = \hat{\psi}_\vartheta(x)\alpha_y^{-1} \ .$$

Moreover, since $\psi(v)$ has no palindromic prefix (resp. suffix) followed (resp. preceded) by y one has

$$\psi(vy) = \psi(v)y\psi(v) \ . \tag{3.10}$$

Thus from (3.9) we obtain

$$\begin{aligned}\hat{\psi}_\vartheta(xvy) &= \phi(\psi(v))\phi(y)\phi(\psi(v))\hat{\psi}_\vartheta(x)\\ &= \phi(\psi(v)y\psi(v))\hat{\psi}_\vartheta(x)\\ &= \phi(\psi(vy))\hat{\psi}_\vartheta(x) \ .\end{aligned}$$

Finally we consider $|v|_y = 0$ and $PAL_\vartheta \cap \mathrm{Pref}(u_0x)y^{-1} = \emptyset$. In this case, since $\hat{\psi}_\vartheta(xv)$ has no ϑ-palindromic suffix preceded by $\bar{y}$ (has no ϑ-palindromic prefix followed by y), we have

$$\hat{\psi}_\vartheta(xvy) = \hat{\psi}_\vartheta(xv)y^{\oplus}\hat{\psi}_\vartheta(xv) \ . \tag{3.11}$$

By induction we then obtain

$$\begin{aligned}\hat{\psi}_\vartheta(xvy) &= \hat{\psi}_\vartheta(xv)y^{\oplus}\hat{\psi}_\vartheta(xv)\\ &= \phi(\psi(v))\hat{\psi}_\vartheta(x)y^{\oplus}\phi(\psi(v))\hat{\psi}_\vartheta(x) \ .\end{aligned} \tag{3.12}$$

In particular, if $v = \varepsilon$,

$$\hat{\psi}_\vartheta(xy) = \hat{\psi}_\vartheta(x)y^{\oplus}\hat{\psi}_\vartheta(x) \ ,$$

so

$$\hat{\psi}_\vartheta(xy)\hat{\psi}_\vartheta(x)^{-1} = \hat{\psi}_\vartheta(x)y^{\oplus} = \phi(y) \ .$$

Then from (3.12) and (3.10) one derives

$$\begin{aligned}\hat{\psi}_{\vartheta}(xvy) &= \phi(\psi(v))\phi(y)\phi(\psi(v))\hat{\psi}_{\vartheta}(x)\\ &= \phi(\psi(v)y\psi(v))\hat{\psi}_{\vartheta}(x)\\ &= \phi(\psi(vy))\hat{\psi}_{\vartheta}(x)\ ,\end{aligned}$$

which completes the proof. □

Example 3.3.5. Let us refer to Example 3.3.1. We have $w = abc$, $u_0 = acbbc$, and ϑ defined by $\bar{a} = b$, $\bar{c} = c$. By the preceding theorem, one has

$$\hat{\psi}_{\vartheta}(abc) = \phi_a(\psi(bc))\hat{\psi}_{\vartheta}(a)\ .$$

Since $\psi(bc) = bcb$, $\phi_a(bcb) = \phi_a(b)\phi_a(c)\phi_a(b)$, and $\hat{\psi}_{\vartheta}(a) = (u_0 a)^{\oplus} = acbbcaacb$, by using (3.5) we obtain

$$\hat{\psi}_{\vartheta}(abc) = acbbcaacbbcaacbcacbbcaacbbcaacb\ ,$$

as already shown in Example 3.3.1.

From Theorem 3.3.4, in the case that w is an infinite word, we obtain:

Theorem 3.3.6. *Let $w \in A^{\omega}$ and $x \in A$. Then*

$$\hat{\psi}_{\vartheta}(xw) = \phi_x(\psi(w))\ ,$$

i. e. , any ϑ-standard word s with seed is the image, by an injective morphism, of the standard Episturmian word whose directive word is obtained by deleting the first letter of the directive word of s.

Proof. Let $w \in A^{\omega}$, $t = \psi(w)$, and $w_n = \mathrm{Pref}(w) \cap A^n$ for all $n \geq 0$. From Theorem 3.3.4, for all $n \geq 0$, $\hat{\psi}_{\vartheta}(xw_n) = \phi(\psi(w_n))\hat{\psi}_{\vartheta}(x)$. Since $\psi(w_{n+1}) = \psi(w_n)\xi_n$ with $\xi_n \in A^+$, one has $\phi(\psi(w_{n+1})) = \phi(\psi(w_n))\phi(\xi_n)$. Hence, $\hat{\psi}_{\vartheta}(xw_{n+1})$ has the same prefix of $\hat{\psi}_{\vartheta}(xw_n)$ of length $|\phi(\psi(w_n))|$, which diverges with n. Since

$$\lim_{n\to\infty} \phi(\psi(w_n)) = \phi\left(\psi(w)\right)\ ,$$

the result follows. □

In the case of an empty seed, from Theorem 3.3.4 one has

$$\psi_\vartheta(xw) = \phi_x(\psi(w))\psi_\vartheta(x) = \phi_x(\psi(w))x^\oplus . \tag{3.13}$$

Moreover, one easily derives that

$$\phi_x(x) = x^\oplus, \quad \phi_x(y) = x^\oplus y^\oplus \text{ for } y \neq x .$$

When $u_0 = \varepsilon$ and $\vartheta = R$, the morphism ϕ_x reduces to μ_x defined as $\mu_x(y) = xy$ for $y \neq x$ and $\mu_x(x) = x$. Since $x^\oplus = x$, from (3.13) one obtains the following formula due to Justin [42]:

$$\psi(xw) = \mu_x(\psi(w))x . \tag{3.14}$$

It is noteworthy that Theorem 3.3.4 provides an alternate proof of Theorem 3.2.2:

Proof of Theorem 3.2.2. It is sufficient to observe that, in the case of an empty seed, $x^\oplus = \mu_\vartheta(x)$ and $\phi_x = \mu_\vartheta \circ \mu_x$, so that by (3.13) and (3.14) one derives:

$$\psi_\vartheta(xw) = (\mu_\vartheta \circ \mu_x)(\psi(w))\mu_\vartheta(x) = \mu_\vartheta(\mu_x(\psi(w))x) = \mu_\vartheta(\psi(xw)) ,$$

as desired. □

Our next goal is to prove a result analogous to Theorem 3.2.7 for words generated by nonempty seeds. However, because of the presence of an arbitrary seed, one cannot hope to prove exactly the same assertion; thus in Theorem 3.3.10 we shall prove that any *sufficiently long* left special factor of a ϑ-standard word with seed is a prefix of it, and give an upper bound for the minimal length from which this occurs, in terms of the length of $(u_0x)^\oplus$.

In the following, we shall set

$$u_1 = \hat{\psi}_\vartheta(x) = (u_0x)^\oplus ,$$

so that $\phi_x(a) = (u_1a)^\oplus u_1^{-1}$ and $|\phi_x(a)| \leq |u_1| + 2$ for any $a \in A$.

For any letter a, u_a will denote (if it exists) the longest ϑ-palindromic suffix (resp. prefix) of u_1 preceded (resp. followed) by $\bar{a}$ (resp. by a). One has then $u_1 = \phi_x(a)u_a$ for any a such that u_a is defined, and $\phi_x(a) = u_1a^\oplus$ otherwise.

Lemma 3.3.7. *Let* $X = \phi_x(A)$. *If* $w \in X^*$, *then* $u_1 \in \mathrm{Pref}(wu_1)$.

Proof. Trivial if $w = \varepsilon$. We shall prove by induction that for all $n \geq 1$, if $w \in X^n$, then $u_1 \in \mathrm{Pref}(wu_1)$. Let $w \in X$. Then there exists $a \in A$ such that $w = \phi(a) = (u_1 a)^{\oplus} u_1^{-1}$. Thus $wu_1 = (u_1 a)^{\oplus}$, so that the statement holds for $n = 1$.

Let us suppose the statement is true for n, we will prove it for $n+1$. If $w \in X^{n+1}$, there exist $a \in A$ and $v \in X^n$ such that $w = \phi(a)v$. By induction, vu_1 can be written as $u_1 v'$ for some $v' \in A^*$. Then one has $wu_1 = \phi(a)u_1 v'$ and, as shown above, u_1 is a prefix of $\phi(a)u_1$, which concludes the proof. □

Recall (cf. [7]) that a pair $(p, q) \in A^* \times A^*$ is *synchronizing* for the code X over the alphabet A if for all $\lambda, \rho \in A^*$,

$$\lambda p q \rho \in X^* \Longrightarrow \lambda p, q\rho \in X^* \ .$$

Proposition 3.3.8. *The pair* (ε, u_1) *is synchronizing for* $X = \phi_x(A)$.

Proof. Since X is a suffix code, it suffices to show that for any $\lambda, \rho \in A^*$,

$$\lambda u_1 \rho \in X^* \Longrightarrow u_1 \rho \in X^* \ .$$

This is trivial if $\lambda = \varepsilon$. Let us factorize $\lambda u_1 \rho$ by the elements of X. Then we can write $\lambda = \lambda' p$ and $u_1 \rho = s\rho'$, where $\lambda', \rho' \in X^*$, and $ps = \phi(a) \in X$ for some letter a (see Figure 3.1). If $p = \varepsilon$, then trivially $u_1 \rho \in X^*$. Suppose then $p \neq \varepsilon$, so that $s \notin X$.

Since $ps \in X$, it follows $|s| \leq |u_1| + 1$. Let us prove that $|s| \leq |u_1|$. By contradiction, suppose $|s| = |u_1| + 1$. Then $\phi(a) = ps = u_1 a \bar{a}$ and $s = u_1 \bar{a}$. Therefore $ps = u_1 a \bar{a} = p u_1 \bar{a}$, so that $u_1 a = p u_1$. This implies $a = p$ and $u_1 = a^k$ for a suitable $k > 0$. Since a is not a ϑ-palindrome, it follows $u_1 \notin PAL_{\vartheta}$, a contradiction.

Thus one has $u_1 = sw$ for some $w \in \mathrm{Pref}(\rho')$. By Lemma 3.3.7, u_1 is a prefix of $\rho' u_1$; clearly, w is a prefix of $\rho' u_1$ too. Therefore w is a prefix of u_1, as $|w| = |u_1| - |s|$. Thus $u_1 = w\bar{s}$, and

$$(u_1 a)^{\oplus} = \phi(a) u_1 = p s u_1 = p s w \bar{s} = p u_1 \bar{s} \ .$$

Since $p \neq \varepsilon$, by Lemma 3.1.7 one obtains $\bar{s} = \varepsilon$. Hence $u_1 \rho = \rho' \in X^*$. □

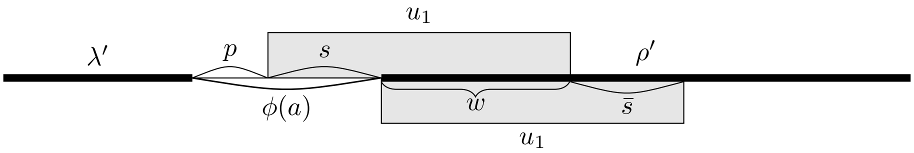

Figure 3.1: Proposition 3.3.8

In the following, if Z is a finite subset of A^*, we shall denote by Z^ω the set of all infinite words which can be factorized by the elements of Z. As is well known (cf. [7]) a word $t \in Z^\omega$ has a unique factorization by means of the elements of Z if and only if Z is a code having *finite deciphering delay*. By Lemma 3.3.7, the code $X = \phi_x(A)$ has the property that there exists an integer $n > 0$ such that $u_1 \in \mathrm{Pref}(v)$ for all $v \in X^n$; from Proposition 3.3.8 it follows that all pairs of $X^n \times X^n$ are synchronizing for X, so that X has a *bounded synchronization delay* and therefore a finite deciphering delay.

Lemma 3.3.9. *Let $X = \phi_x(A)$ and $w = ru_1azs \in X^*$, with $a, z \in A$ and $r, s \in A^*$. If we set $v' = \phi_x(a)^{-1}u_1az$, then $(r, v's)$ is in $X^* \times X^*$ and it is an occurrence of $\phi_x(a)$ in w.*

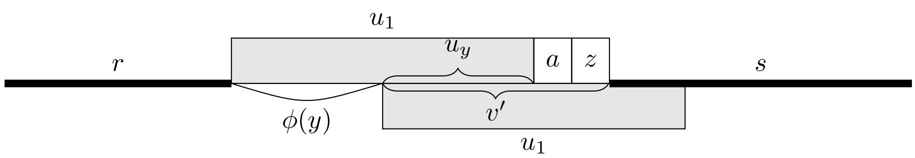

Figure 3.2: Lemma 3.3.9

Proof. Let $w \in X^*$ be such that $w = ru_1azs$, with $z \in A$. From Proposition 3.3.8 we have that r and u_1azs are in X^*. Let $y \in A$ be a letter such that $v = \phi(y)^{-1}u_1azs$ is in X^* and set $v' = \phi(y)^{-1}u_1az$. It is clear from the definition of ϕ that either $v' = \varepsilon$, $v' = z$ or $v' = u_yaz$, where u_y is the longest ϑ-palindromic suffix of u_1 preceded by $\bar{y}$. In the first two cases, it must be $\phi(y) = u_1a^{\oplus}$, so that $a = y$; let then $v' = u_yaz$ (see Figure 3.2). Since $v = v's \in X^*$, from Lemma 3.3.7 it follows that u_1 is a prefix of $v'su_1$, so u_ya, whose length is less than $|u_1|$, is a prefix of u_1. By definition, u_y is a prefix of

u_1 followed by y, hence $u_y y = u_y a$ and $a = y$. Thus $(r, v's) \in X^* \times X^*$ is an occurrence of $\phi(a)$ in w. $\square$

Theorem 3.3.10. *Let $t = \hat{\psi}_\vartheta(x\Delta)$ be a ϑ-standard word with seed. Then there exists an integer $N \geq 0$ such that any left special factor of t of length greater than or equal to N is a prefix of t.*

Proof. Set $z = \psi(\Delta) = z_1 z_2 \cdots z_n \cdots$, where $z_i \in A$ for all $i \geq 1$. From Theorem 3.3.4 we have that $t = \phi(z)$, so that t can be factorized uniquely as

$$t = \phi(z_1)\phi(z_2)\cdots\phi(z_n)\cdots \in X^\omega ,$$

where $X = \phi_x(A)$. We shall prove that each left special factor w of t longer than $2|u_1|+2$ is also a prefix of t. Since w is left special, there exist two different occurrences of w in t preceded by distinct letters, say a and b. Moreover, since $|w| > 2|u_1| + 2$, we can write

$$w = p\phi(z_{i+1}\cdots z_{i+h})s = p'\phi(z_{j+1}\cdots z_{j+k})s' , \tag{3.15}$$

where $\phi(z_i) = rap$, $\phi(z_j) = r'bp'$, $\phi(z_{i+h+1}) = s\lambda$, and $\phi(z_{j+k+1}) = s'\lambda'$, with $\lambda, \lambda' \in A^+$ and i, j, h, k positive integers. Thus one can rewrite t as

$$t = \phi(z_1\cdots z_{i-1})raw\lambda\phi(z_{i+h+2}\cdots) = \phi(z_1\cdots z_{j-1})r'bw\lambda'\phi(z_{j+k+2}\cdots).$$

Without loss of generality, we can suppose $|p| \leq |p'|$. From (3.15) and from the preceding equation, we have

$$rap'\phi(z_{j+1}\cdots z_{j+k})s'\lambda\phi(z_{i+h+2}\cdots) \in X^\omega.$$

Since $|w| > 2|u_1| + 2$ and $p' \leq |u_1| + 1$, one has $|\phi(z_{j+1}\cdots z_{j+k})s'| > |u_1| + 1$, so that from Lemma 3.3.7, u_1 is a prefix of $\phi(z_{j+1}\cdots z_{j+k})s'\lambda' u_1$ and then of $\phi(z_{j+1}\cdots z_{j+k})s'$.

By Proposition 3.3.8, $(p', \phi(z_{j+1}\cdots z_{j+k})s')$ is a synchronizing pair for X, so that rap' is in X^*. If $p' \neq \varepsilon$, then $r'bp'$ is the only word of the code X having p' as a suffix (recall that any codeword of X is determined by its last letter); hence it should be a suffix of rap', which is clearly a contradiction as $a \neq b$. Then $p' = \varepsilon$, that implies also $p = \varepsilon$. Thus, we can write

$$t = \phi(z_1\cdots z_i)w\lambda\phi(z_{i+h+2}\cdots) = \phi(z_1\cdots z_j)w\lambda'\phi(z_{j+k+2}\cdots),$$

and $z_i \neq z_j$, as w is left special. Since

$$w = \phi(z_{i+1} \cdots z_{i+h})s = \phi(z_{j+1} \cdots z_{j+k})s'$$

is longer than $2|u_1| + 2$, and $|s|, |s'| \leq |u_1| + 1$, there exists a letter $c \in A$ such that $u_1 c$ is a prefix of both $\phi(z_{i+1} \cdots z_{i+h})$ and $\phi(z_{j+1} \cdots z_{j+k})$. By Lemma 3.3.9 one has $\phi(z_{i+1} \cdots z_{i+h}) = \phi(c)\rho$ and $\phi(z_{j+1} \cdots z_{j+k}) = \phi(c)\rho'$ with $\rho, \rho' \in X^*$, so that $z_{i+1} = z_{j+1} = c$ since X is a code.

Let l be the greatest integer such that $z_{i+m} = z_{j+m}$ for all $m \leq l$. Then both $z_i z_{i+1} \cdots z_{i+l}$ and $z_j z_{j+1} \cdots z_{j+l} = z_j z_{i+1} \cdots z_{i+l}$ are factors of z. Since $z_i \neq z_j$, $z_{i+1} \cdots z_{i+l}$ is a left special factor of the Episturmian word z, thus a prefix of z, i. e. , $z_{i+1} \cdots z_{i+l} = z_1 \cdots z_l$. Hence $\phi(z_{i+1} \cdots z_{i+l})$ is a prefix of t.

Now let us suppose that $w' = \phi(z_{i+l+1} \cdots z_{i+h})s = \phi(z_{j+l+1} \cdots z_{j+k})s'$ is strictly longer than u_1. By Lemma 3.3.7, there exists a letter d such that $u_1 d$ is a prefix of w', so, by applying Lemma 3.3.9 to $w'\lambda \in X^*$ and to $w'\lambda' \in X^*$ one derives $\phi(z_{i+l+1}) = \phi(z_{j+l+1}) = \phi(d)$, contradicting the fact that $i + l$ was the largest of such indexes. Then $|w'| \leq |u_1|$. By Lemma 3.3.7, u_1 is a prefix of $w'\lambda u_1$. Thus w' is a prefix of u_1 and $w = \phi(z_{i+1} \cdots z_{i+l})w'$ is a prefix of $\phi(z_{i+1} \cdots z_{i+l})u_1 = \phi(z_1 \cdots z_l)u_1$. Let m be an integer such that $|u_1| \leq |\phi(z_{l+1} \cdots z_{l+m})|$. By Lemma 3.3.7, u_1 is a prefix of $\phi(z_{l+1} \cdots z_{l+m})$ and $\phi(z_1 \cdots z_l)u_1$ is a prefix of $\phi(z_1 \cdots z_{l+m})$ which is a prefix of t. In conclusion, we obtain that w is a prefix of t. $\square$

We observe that the proof of the preceding theorem shows that for a ϑ-standard word s with seed u_0, all left special factors of length greater than or equal to $N = 2|u_1| + 3$ are prefixes of s. However, this bound is not tight. In fact, for instance, if $u_0 = \varepsilon$ then $N = 5$, whereas from Theorem 3.2.7 one has that all left special factors of a ϑ-standard word s, having length at least 3, are prefixes of s.

The following lemma, whose proof is in [27], will be useful in the sequel.

Lemma 3.3.11. *Let $u \in A^*$ and $w = (ux)^{\oplus}$, where $x \in A$. If p is any prefix of w of length $|p| > |u|$, then $p^{\oplus} = w$.*

Proposition 3.3.12. *Let $s = \hat{\psi}_{\vartheta}(\Delta)$ be a ϑ-standard word with a seed u_0 of length k. The following hold:*

1. *A word w with $|w| > k$ is a prefix of s if and only if $w^{\oplus}$ is a prefix of s,*

2. *the set of all ϑ-palindromic prefixes of s is given by*

$$\hat{\psi}_{\vartheta}(\mathrm{Pref}(\Delta) \setminus \{\varepsilon\}) \cup (PAL_{\vartheta} \cap \mathrm{Pref}(u_0)) \, , \tag{3.16}$$

3. *s is closed under ϑ.*

Proof. If $w^{\oplus}$ is a prefix of s, then trivially w is a prefix of s. Conversely, suppose that w is a prefix of s with $|w| > k$. If $\Delta = xt_1t_2 \cdots t_n \cdots$ with $x \in A$ and $t_i \in A$, $i > 0$. Let us set $u_1 = (u_0x)^{\oplus} = \hat{\psi}_{\vartheta}(x)$ and for $n > 1$, $u_{n+1} = \hat{\psi}_{\vartheta}(xt_1 \cdots t_n)$, so that $u_{n+1} = (u_nt_n)^{\oplus}$. We consider the least n such that $|u_n| < |w| \leq |u_{n+1}|$. By Lemma 3.3.11 one has $w^{\oplus} = u_{n+1} \in \mathrm{Pref}(s)$. This proves point 1.

By the definition of ϑ-standard words with seed, all the words in the set (3.16) are ϑ-palindromic prefixes of s. Conversely, let w be a ϑ-palindromic prefix of s. If $|w| \leq k$, then trivially $w \in PAL_{\vartheta} \cap \mathrm{Pref}(u_0)$. If $|w| > k$, then by following the same argument used for point 1, one has that there exists an integer $n > 0$ such that $w = w^{\oplus} = u_n \in \hat{\psi}_{\vartheta}(\mathrm{Pref}(\Delta))$. This proves point 2.

Let w be a factor of s. Since there are infinitely many ϑ-palindromic prefixes of s, there exists a ϑ-palindromic prefix u having w as a factor. Therefore, also $\bar{w}$ is a factor of u and of s. This concludes the proof. ☐

By a generalization of an argument used in [29] for Episturmian words, one can prove the following:

Proposition 3.3.13. *Any ϑ-standard word s with seed is uniformly recurrent.*

Proof. Let $\Delta(s) = xt_1 \cdots t_n \cdots$ be the directive word of $s = \lim_{n\to\infty} u_n$, where $u_1 = (u_0x)^{\oplus}$ and $u_{n+1} = (u_nt_n)^{\oplus}$ for $n > 0$. The word s is trivially recurrent. We shall prove that the shifts of the first returns to any factor v of s are bounded by a constant. Let m be the smallest integer such that $v \in \mathrm{Fact}(u_m)$. Let us set $p = u_m$ and let ρ_n be the maximal shift of all first returns to p in u_n, for all $n > m$. Since $u_{n+1} = (u_nt_n)^{\oplus}$, one has $|u_{n+1}| \leq 2|u_n| + 2$, where such upper bound is reached if and only if $u_{n+1} = u_nt_n\bar{t}_nu_n$. This implies that

$\rho_{m+1} \leq |p| + 2$. Moreover, for all $n > m$ we have $\rho_{n+1} \leq \max\{\rho_n, |p| + 2\}$. Indeed, let w be a first return to p in u_{n+1} of maximal length, so that its shift is ρ_{n+1}. If $w \in \mathrm{Fact}(u_n)$, then $\rho_{n+1} = \rho_n$. Let us suppose that w is not a factor of u_n. We set $u_n = \lambda p = p\bar{\lambda}$ and $u_{n+1} = \alpha w \beta$ with $\alpha, \beta, \lambda \in A^*$. Then $|\alpha| \geq |\lambda|$ and $|\beta| \geq |\lambda|$, otherwise w would be a factor of u_n. Therefore, as $|u_{n+1}| \leq 2|u_n| + 2$, we obtain

$$|w| \leq |u_{n+1}| - 2|\lambda| = |u_{n+1}| - 2|u_n| + 2|p| \leq 2|p| + 2 \ ,$$

so that $\rho_{n+1} \leq |p| + 2$. Thus in any case $\rho_{n+1} \leq \max\{\rho_n, |p| + 2\}$. As $\rho_{m+1} \leq |p| + 2$, it follows that $\rho_n \leq |p| + 2$ for all $n > m$.

Since v is a factor of u_m, the shifts of all first returns of v in s are upper limited by $|p| + 2 = |u_m| + 2$. □

Let $\hat{\psi}_\vartheta(\Delta)$ be a ϑ-standard word with seed u_0 and directive word $\Delta = xt_1t_2\cdots t_n\cdots$. Define the endomorphism ϕ_x of A^* by setting

$$\phi_x(a) = \hat{\psi}_\vartheta(xa)\hat{\psi}_\vartheta(x)^{-1}$$

for any letter $a \in A$. From the definition, one has that ϕ_x depends on ϑ and u_0; moreover, $\phi_x(a)$ ends with $\bar{a}$ for all $a \in A$, so that any word of the set $X = \phi_x(A)$ is uniquely determined by its last letter. Thus X is a suffix code and ϕ_x is an injective morphism.

Example 3.3.14. Let A, ϑ, and u_0 be defined as in Example 3.3.2, and let $x = a$. Then

$$\begin{aligned}
\phi_a(a) &= \hat{\psi}_\vartheta(aa)\hat{\psi}_\vartheta(a)^{-1} = abaca \ , \\
\phi_a(b) &= \hat{\psi}_\vartheta(ab)\hat{\psi}_\vartheta(a)^{-1} = abac \ , \\
\phi_a(c) &= \hat{\psi}_\vartheta(ac)\hat{\psi}_\vartheta(a)^{-1} = abacacb \ .
\end{aligned}$$

The following important theorem on ϑ-standard words with seed, whose proof is in [13], shows that such words are morphic images of standard Episturmian words.

Theorem 3.3.15. *Let $w \in A^\omega$ and $x \in A$. Then*

$$\hat{\psi}_\vartheta(xw) = \phi_x(\psi(w)) \ ,$$

i.e., any ϑ-standard word s with seed is the image, by an injective morphism, of the standard Episturmian word whose directive word is obtained by deleting the first letter of the directive word of s.

Proposition 3.3.16. *If s is a ϑ-standard word with seed and two letters of A occur infinitely often in $\Delta(s)$, then any prefix of s is a left special factor of s.*

Proof. A prefix p of s is also a prefix of any ϑ-palindromic prefix B of s such that $|p| \leq |B|$. Since there exist two distinct letters, say a and b, which occur infinitely often in $\Delta(s)$, one has $Ba, Bb \in \mathrm{Fact}(s)$. Therefore, $\bar{p}a, \bar{p}b \in \mathrm{Fact}(s)$, i.e., $\bar{p}$ is right special. Since by Proposition 3.2.3, s is closed under ϑ, one has $\bar{a}p, \bar{b}p \in \mathrm{Fact}(s)$; as $\bar{a} \neq \bar{b}$, p is left special. $\square$

In general, a ϑ-standard word with seed (empty or not) can have left special factors which are not prefixes. However, Theorem 3.3.10, shows that all sufficiently long left special factors of a ϑ-standard word with seed are prefixes of it. One of the main results of Section 3.5 shows that the previous property on left special factors, along with closure under ϑ, characterizes ϑ-standard words with seed.

An infinite word $s \in A^\omega$ is called a *ϑ-word with seed* if there exists a ϑ-standard word t with seed such that $\mathrm{Fact}(s) = \mathrm{Fact}(t)$.

3.4 ϑ-Episturmian words

In [12] *standard ϑ-Episturmian* words were naturally defined by substituting, in the definition of standard Episturmian words, the closure under reversal with the *closure under ϑ*. Thus an infinite word s is standard ϑ-Episturmian if it satisfies the following two conditions:

1. for any $w \in \mathrm{Fact}\, s$, one has $\bar{w} \in \mathrm{Fact}\, s$,

2. for any left special factor w of s, one has $w \in \mathrm{Pref}\, s$.

We denote by $SEpi_\vartheta$ the set of all standard ϑ-Episturmian words on the alphabet A.

More generally, it will be useful to introduce for any $N \geq 0$ the family $SW_\vartheta(N)$ of all infinite words w which are closed under ϑ and such that every left special factor of w whose length is at least N is a prefix of w. Moreover, by $W_\vartheta(N)$ we denote the class of all infinite words having the same set of factors as some word in $SW_\vartheta(N)$. Thus $SW_\vartheta(0) = SEpi_\vartheta$ and $W_\vartheta(0) = Epi_\vartheta$. By Theorem 3.2.7, the class of ϑ-standard words is included in $SW_\vartheta(3)$.

Proposition 3.4.1. *An infinite word s is in $W_\vartheta(N)$ if and only if s is closed under ϑ and it has at most one left special factor of any length greater than or equal to N.*

Proof. The "only if" part follows immediately from the fact that $\mathrm{Fact}(s) = \mathrm{Fact}(t)$ for some $t \in SW_\vartheta(N)$. Let us prove the "if" part. Let us first suppose that s has infinitely many left special factors. Hence s has exactly one left special factor for each length $n \geq N$, say v_n. Then for any $n \geq N$, v_n is a prefix of v_{n+1}, so that

$$t = \lim_{n\to\infty} v_n$$

is a well-defined infinite word. Trivially $\mathrm{Fact}(t) \subseteq \mathrm{Fact}(s)$; thus to prove that $\mathrm{Fact}(t) = \mathrm{Fact}(s)$ it suffices to show that any given factor w of s with $|w| \geq N$ is a factor of some v_n, $n \geq N$. Since s is closed under ϑ, $\bar{w}$ is a factor of s. Let p be a prefix of s ending in $\bar{w}$. Since s is recurrent, we can consider a prefix of s of the kind pup for some $u \in A^*$. Then there exists $v \in A^*$ such that pv is a right special factor of s, for otherwise one would have $s = (pu)^\omega$, contradicting the fact that s has infinitely many left special factors. Hence $\bar{w}v$ is a right special factor of s, so that $\bar{v}w$ is a left special factor of s. Since $|w| \geq N$, we have $|\bar{v}w| \geq N$ and therefore $\bar{v}w \in \mathrm{Pref}(t)$; thus $\mathrm{Fact}(t) = \mathrm{Fact}(s)$ as desired. This implies that any left special factor of t is also left special in s. It follows that $t \in SW_\vartheta(N)$.

Now suppose that s has only finitely many left special factors. As is well known, this implies that s is eventually periodic, and hence periodic since it is recurrent. Let then w be the longest left special factor of s, and let $s = \lambda w s'$ for some $\lambda \in A^*$ and $s' \in A^\omega$. Then $t = ws'$ has the same set of factors as s. This implies that t is a word of $SW_\vartheta(N)$. $\square$

As an immediate consequence, one obtains:

Corollary 3.4.2. *An infinite word is ϑ-Episturmian if and only if it is closed under ϑ and it has at most one left special factor of each length.*

Remark. In the case of a binary alphabet $A = \{a, b\}$, by definition any word $s \in Epi_\vartheta$ has a subword complexity λ_s such that $\lambda_s(n) \leq n + 1$ for all $n \geq 0$. It follows that any word in Epi_ϑ is either Sturmian or periodic. In particular, if $\vartheta = E \circ R$, then the word s cannot be Sturmian, since any Sturmian word has either aa or bb as a factor, but not both, whereas s, being closed under ϑ, does not satisfy this requirement. Thus Epi_ϑ contains only the two periodic words $(ab)^\omega$ and $(ba)^\omega$, whereas Epi_R contains all Sturmian words.

The following two propositions, proved in [12], give methods for constructing standard ϑ-Episturmian words.

Proposition 3.4.3. *Let s be a ϑ-standard word over A, and $B = \mathrm{alph}(\Delta(s))$. Then s is standard ϑ-Episturmian if and only if*

$$x \in B,\, x \neq \bar{x} \implies \bar{x} \notin B\ .$$

Example 3.4.4. Let $A = \{a, b, c, d, e\}$, $\Delta = (acd)^\omega$, and ϑ be defined by $\bar{a} = b$, $\bar{c} = c$, and $\bar{d} = e$. The ϑ-standard word $\psi_\vartheta(\Delta) = abcabdeabcaba \cdots$ is standard ϑ-Episturmian.

Proposition 3.4.5 (Proposition 3.6.7). *Let $\phi : X^* \to A^*$ be a nonerasing morphism such that*

1. *$\phi(x) \in PAL_\vartheta$ for all $x \in X$,*

2. *$\mathrm{alph}\,\phi(x) \cap \mathrm{alph}\,\phi(y) = \emptyset$ if $x, y \in X$ and $x \neq y$,*

3. *3 $|\phi(x)|_a \leq 1$ for all $x \in X$ and $a \in A$.*

Then for any standard Episturmian word $t \in X^\omega$, $s = \phi(t)$ is a standard ϑ-Episturmian word.

Example 3.4.6. Let $A = \{a, b, c, d, e\}$, $\bar{a} = b$, $\bar{c} = c$, $\bar{d} = e$, $X = \{x, y\}$, and $s = g(t)$, where $t = xxyxxxyxxxyxxy \cdots \in SEpi(X)$, $\Delta(t) = (xxy)^\omega$, $g(x) = acb$, and $g(y) = de$, so that

$$s = acbacbdeacbacbacbde \cdots\ . \tag{3.17}$$

By the previous proposition, the word s is standard ϑ-Episturmian, but it is not ϑ-standard, as $a^\oplus = ab \notin \mathrm{Pref}\, s$.

It is easy to prove (see [12]) that every standard ϑ-Episturmian word has infinitely many ϑ-palindromic prefixes. This implies, by Proposition 3.6.2, the following:

Proposition 3.4.7. *Every standard ϑ-Episturmian word s admits a (unique) factorization by the elements of $\mathcal{P}_\vartheta$, that is,*

$$s = \pi_1 \pi_2 \cdots \pi_n \cdots ,$$

where $\pi_i \in \mathcal{P}_\vartheta$ for $i \geq 1$.

For a given standard ϑ-Episturmian word s, such factorization will be called *canonical* in the sequel. For instance, in the case of the standard ϑ-Episturmian word of Example 3.4.6, the canonical factorization is:

$$acb \cdot acb \cdot de \cdot acb \cdot acb \cdot acb \cdot de \cdots .$$

The following important lemma was proved in [12]:

Lemma 3.4.8 (Theorem 3.6.4). *Let s be a standard ϑ-Episturmian word, and $s = \pi_1 \cdots \pi_n \cdots$ be its canonical factorization. For all $i \geq 1$, any proper and nonempty prefix of π_i is not right special in s.*

In the following, for a given standard ϑ-Episturmian word s we shall denote by

$$\Pi_s = \{\pi_n \mid n \geq 1\} \tag{3.18}$$

the set of words of $\mathcal{P}_\vartheta$ appearing in its canonical factorization $s = \pi_1 \pi_2 \cdots$.

Theorem 3.4.9. *Let $s \in SEpi_\vartheta$. Then Π_s is a normal code.*

Proof. Any nonempty prefix p of a word of Π_s does not belong to Π_s, since Π_s is a biprefix code. Moreover, $p \notin RS\,\Pi_s$ as otherwise it would be a right special factor of s, and this is excluded by Lemma 3.6.4. Hence Π_s is a right normal code. Since s is closed under ϑ and $\Pi_s \subseteq PAL_\vartheta$, it follows that Π_s is also left normal. □

The following result shows that no two words of Π_s overlap properly.

Theorem 3.4.10. *Let $s \in SEpi_\vartheta$. Then Π_s is an overlap-free code.*

Proof. If $\text{card}\Pi_s = 1$ the statement is trivial since an element of $\mathcal{P}_\vartheta$ cannot overlap properly with itself as it is unbordered. Let then $\pi, \pi' \in \Pi_s$ be such that $\pi \neq \pi'$. By contradiction, let us suppose that there exists a nonempty $u \in \text{Suff}\,\pi \cap \text{Pref}\,\pi'$ (which we can assume without loss of generality, since it occurs if and only if $\bar{u} \in \text{Suff}\,\pi' \cap \text{Pref}\,\pi$). We have $|\pi| \geq 2|u|$ and $|\pi'| \geq 2|u|$, for otherwise u would overlap properly with $\bar{u}$ and so it would have a nonempty ϑ-palindromic prefix (or suffix), which is absurd. Then there exist $v, v' \in PAL_\vartheta$ such that $\pi = \bar{u}vu$ and $\pi' = uv'\bar{u}$.

Without loss of generality, we can assume that π occurs before π' in the canonical factorization of s, so that there exists $\lambda \in (\Pi_s \setminus \{\pi'\})^*$ such that $\lambda\pi \in \text{Pref}\,s$. Since by Lemma 3.6.4 any proper prefix of π cannot be right special in s, each occurrence of $\bar{u}$ must be followed by vu; the same argument applies to π', so each occurrence of u in s must be followed by $v'\bar{u}$. Therefore we have

$$s = \lambda(\bar{u}vuv')^\omega = \lambda(\pi v')^\omega \ .$$

As v' is a ϑ-palindromic proper factor of π', it must be in $(\mathcal{P}_\vartheta \setminus \{\pi'\})^*$, as well as $\pi v'$ and, by definition, λ. Thus we have obtained that $s \in (\Pi_s \setminus \{\pi'\})^\omega$, and so $\pi' \notin \Pi_s$, which is clearly a contradiction. Then π and π' cannot overlap properly. ☐

The following theorem, proved in [12, Theorem 5.5], shows, in particular, that any standard ϑ-Episturmian word is a morphic image, by a suitable injective morphism, of a standard Episturmian word. We report here a direct proof based on the previous results.

Theorem 3.4.11. *Let s be a standard ϑ-Episturmian word. Then $f(s)$ is a standard Episturmian word, and the restriction of f to Π_s is injective, i.e., if π_i and π_j occur in the factorization of s over $\mathcal{P}_\vartheta$, and $\pi_i^f = \pi_j^f$, then $\pi_i = \pi_j$.*

Proof. Since $s \in SEpi_\vartheta$, by Theorems 3.4.9 and 3.4.10 the code Π_s is biprefix, overlap-free, and normal. By Proposition 1.3.1, the restriction to Π_s of the map f defined by (3.2) is injective. Let $B = f(\Pi_s) \subseteq A$ and denote by $g : B^* \to A^*$ the injective morphism defined by $g(\pi^f) = \pi$ for any $\pi^f \in B$. One has $s = g(t)$ for some $t \in B^\omega$. Let us now show that $t \in SEpi(B)$. Indeed, since s has infinitely many ϑ-palindromic prefixes, by Proposition 3.1.9 it follows that t has infinitely many palindromic prefixes, so that it is closed under reversal. Let now w be a left special factor of t, and let $a, b \in B$, $a \neq b$, be such that $aw, bw \in \operatorname{Fact} t$. Thus $g(a)g(w), g(b)g(w) \in \operatorname{Fact} s$. Since $g(a)^f \neq g(b)^f$, we have $g(a)^\ell \neq g(b)^\ell$, so that $g(w)$ is a left special factor of s, and then a prefix of it. From Lemma 1.3.2 it follows $w \in \operatorname{Pref} t$. □

3.5 Connection between classes of generalized Episturmian words

Let us recall that $SW_\vartheta(N)$ is the family of all infinite words w which are closed under ϑ and such that every left special factor of w whose length is at least N is a prefix of w. Trivially, we have $SW_\vartheta(N) \subseteq SW_\vartheta(N+1)$. Let us denote by SW_ϑ the class of words which are in $SW_\vartheta(N)$ for some $N \geq 0$, i.e.,

$$SW_\vartheta = \bigcup_{N \geq 0} SW_\vartheta(N) \, .$$

One of the main results is the proof that SW_ϑ coincides with the class of ϑ-standard words with seed (cf. Theorem 3.5.4). As a corollary, we will derive that any standard ϑ-Episturmian word is a ϑ-standard word with seed.

For the sake of clarity, we report in Table 3.1 the definitions and the notations of the different classes of words introduced so far. We consider only the standard case, since the "non-standard" words of a given class are defined by the property of having the same set of factors as a standard one.

In order to prove the main theorem, we need some preliminary results.

Lemma 3.5.1. *Let $w \in SW_\vartheta(N)$ and u be a ϑ-palindromic factor of w such that $|u| \geq N$. Then the leftmost occurrence of u in w is a median factor of a ϑ-palindromic prefix of w.*

Proof. By contradiction, suppose that $w = \lambda x v u \bar{v} \bar{y} w'$, for some letters $x, y \in A$ with $x \neq y$, and words $\lambda, v \in A^*$, $w' \in A^\omega$. Since w is closed under ϑ, both $xvu\bar{v}$ and $yvu\bar{v}$ are factors of w, so that $vu\bar{v}$ is a left special factor of w of length $|vu\bar{v}| \geq N$, and hence a prefix of it. This leads to a contradiction, because we have found an occurrence of u in w before the leftmost one. □

Table 3.1: Summary of the generalizations of standard Episturmian words

Name	Symbol	Definition
ϑ-standard with seed	SW_ϑ [a]	Generated by iterated ϑ-palindrome closure, starting from any seed
ϑ-standard		Generated by iterated ϑ-palindrome closure, starting from ε
Standard ϑ-Episturmian	$SEpi_\vartheta = SW_\vartheta(0)$	Closed under ϑ, and all left special factors are prefixes
	$SW_\vartheta(N)$	Closed under ϑ, and all left special factors of length at least N are prefixes

[a] After Theorem 3.5.4

Proposition 3.5.2. *Any word in SW_ϑ has infinitely many ϑ-palindromic prefixes.*

Proof. Let $w \in SW_\vartheta(N)$ for a suitable $N \geq 0$, and u be a prefix of w, with $|u| \geq N$. We shall prove that w has a ϑ-palindromic prefix whose length is at least $|u|$, from which the assertion will follow.

Let $\alpha\bar{u}$ ($\alpha \in A^*$) be the prefix of w ending with the first occurrence of $\bar{u}$. Since u is a prefix of w, one has $\alpha\bar{u} = u\beta$ for a suitable $\beta \in A^*$. If $\beta = \varepsilon$, then $\alpha = \varepsilon$ and $u = \bar{u}$, so that $\alpha\bar{u} = u$ is the desired ϑ-palindromic prefix.

Then suppose $\beta = x_1 x_2 \cdots x_n$ with $x_i \in A$ for $i = 1, \ldots, n$. As $|\alpha| = |\beta|$, one has $\alpha = y_n \ldots y_1$ for some $y_i \in A$, $i = 1, \ldots, n$. Since $\alpha \neq \varepsilon$, one has $u \neq \bar{u}$, so that $\bar{u}$ is not left special in w. Hence $y_1\bar{u}$ is the only left extension of $\bar{u}$ in w. As w is closed under ϑ, $u\bar{y}_1$ is the only right extension of u in w. This implies $y_1 = \bar{x}_1$.

Since $\alpha\bar{u} = y_n \cdots y_2 \bar{x}_1 \bar{u}$ ends with the first occurrence of $\bar{u}$ (and hence with the first occurrence of $\bar{x}_1\bar{u}$), one can apply the same argument as above to the prefix ux_1, in order to show that $y_2 = \bar{x}_2$. Continuing this way, one eventually obtains $y_i = \bar{x}_i$ for all $i = 1, \ldots, n$, so that $\alpha = \bar{\beta}$ and $\alpha\bar{u}$ is again the desired

ϑ-palindromic prefix of w. □

For a (fixed but arbitrary) word $w \in SW_\vartheta$ we denote by $(B_n)_{n\geq 1}$ the sequence of all ϑ-palindromic prefixes of w, ordered by increasing length. Moreover, for any $i > 0$ let x_i be the unique letter such that $B_i x_i$ is a prefix of w. The infinite word $x = x_1 x_2 \cdots x_n \cdots$ will be called the *subdirective word* of w. The proof of Proposition 3.5.2 shows that for any $i > 0$, B_{i+1} coincides with the prefix of w ending with the first occurrence of $\bar{x}_i B_i$.

The next lemma shows that, under suitable circumstances, a stronger relation holds.

Lemma 3.5.3. *Let $w \in SW_\vartheta(N)$. With the above notation, let $n > 1$ be such that $x_n = x_k$ for some $k < n$ with $|B_k| \geq N-2$. Then $B_{n+1} = (B_n x_n)^\oplus$.*

Proof. Let k be the greatest integer satisfying the hypotheses of the lemma. Let us first prove that $Q = \bar{x}_n B_k x_n$ does not occur in B_n. By contradiction, consider the rightmost occurrence of Q in B_n, i.e., let $Q\rho$ be a suffix of B_n such that Q does not occur in any shorter suffix. If $|\rho| \leq |B_k|$, then one can easily show that the suffix $Q\rho x_n$ of $B_n x_n$ is a ϑ-palindrome, which is absurd because its length is $|Q\rho x_n| > |Q|$.

Suppose then $Q\rho = \bar{x}_n B_k x_n v \bar{x}_n B_k$ for some $v \in A^*$. Since $Q\rho$ is a suffix of B_n, one has that $\bar{\rho} Q = B_k x_n \bar{v} Q$ is a prefix of B_n (see Figure 3.3). Now there is no proper suffix u of $\bar{v}$ such that uQ is left special in w. Indeed, if such u existed, then uQ would be a prefix of B_n, and so $Q\bar{u}$ would be a suffix of B_n, contradicting (as $|u| < |\rho|$) the fact that $Q\rho$ begins with the rightmost occurrence of Q in B_n. Hence every occurrence of Q in w is preceded by $\bar{v}$. Since $\rho x_n = v \bar{x}_n B_k x_n$ is a factor of w, one obtains $v = \bar{v}$, so that $Q\rho x_n = \bar{x}_n B_k x_n v \bar{x}_n B_k x_n$ is a ϑ-palindromic suffix of $B_n x_n$ longer than Q, a contradiction.

Thus Q does not occur in B_n. Since Q is the longest ϑ-palindromic suffix of $B_n x_n$, we can write

$$w = B_n x_n w' = sQw' \ ,$$

where (s, w') is the leftmost occurrence of Q in w. By Lemma 3.5.1, $sQ\bar{s} = (B_n x_n)^\oplus$ is a prefix of w. From this one derives $B_{n+1} = (B_n x_n)^\oplus$. □

Theorem 3.5.4. *Let $s \in A^\omega$. The following conditions are equivalent:*

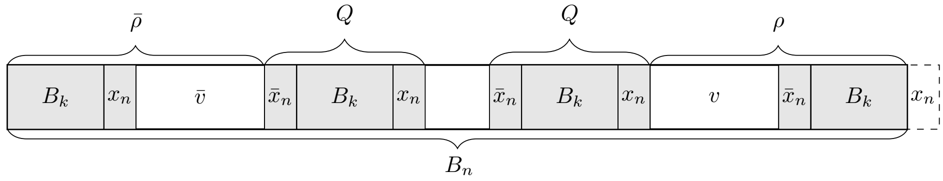

Figure 3.3: Lemma 3.5.3

1. *$s \in SW_\vartheta$,*

2. *s has infinitely many ϑ-palindromic prefixes, and if $(B_n)_{n>0}$ is the sequence of all its ϑ-palindromic prefixes ordered by increasing length, there exists an integer h such that*

$$B_{n+1} = (B_n x_n)^{\oplus} \ ,$$

for all $n \geq h$, for a suitable letter x_n,

3. *s is a ϑ-standard word with seed.*

Proof. *1.⇒2.* Let $s \in SW_\vartheta(N)$, $x_1 x_2 \cdots x_n \cdots$ be its subdirective word, and $(B_i)_{i>0}$ the sequence of all ϑ-palindromic prefixes of s. We consider the minimal integer p such that $|B_p| \geq N-2$. We set $x_{[p]} = x_p x_{p+1} \cdots x_n \cdots \in A^\omega$, and take the minimal m such that $\mathrm{alph}(x_p \cdots x_{p+m}) = \mathrm{alph}(x_{[p]})$. Let $h = p + m + 1$. Then for all $n \geq h$, there exists k with $p \leq k \leq p+m$ such that $x_k = x_n$. Since $k \geq p$ one has $|B_k| \geq N-2$, so that by Lemma 3.5.3, $B_{n+1} = (B_n x_n)^{\oplus}$.

2.⇒3. Let $\hat{\psi}_\vartheta(\Delta)$ be the ϑ-standard word with seed $u_0 = B_h$ and directive word $\Delta = x_h x_{h+1} \cdots x_n \cdots$. One has then $\hat{\psi}_\vartheta(\Delta) = s$.

3.⇒1. This follows from Theorem 3.3.10. □

Let us set

$$W_\vartheta = \bigcup_{N \geq 0} W_\vartheta(N) \ .$$

The following corollary is a straightforward consequence of the preceding theorem.

Corollary 3.5.5. *W_ϑ coincides with the set of all ϑ-words with seed.*

Let $s \in SW_\vartheta(N)$. We call *critical integer* h of s the minimal integer p with the property that for all $n \geq p$ there exists $k < n$ such that $|B_k| \geq N-2$ and $x_n = x_k$. We observe that the proof of Theorem 3.5.4 shows that for any given $s \in SW_\vartheta(N)$ having critical integer h, one has that for all $n \geq h$, $B_{n+1} = (B_n x_n)^\oplus$.

Corollary 3.5.6. *Any standard ϑ-Episturmian word is a ϑ-standard word with seed. Moreover, if $s \in SEpi_\vartheta$ and $x = x_1x_2\cdots x_n\cdots$ is its subdirective word, then the critical integer h of s is equal to the minimal integer p such that* $\mathrm{alph}(x) = \mathrm{alph}(x_1 \cdots x_{p-1})$.

Proof. It is sufficient to observe that a standard ϑ-Episturmian word s is in $SW_\vartheta(0)$ because all its left special factors are prefixes of s. Therefore by Theorem 3.5.4, s is a ϑ-standard word with seed B_h. Since for all $n > 0$ one has $|B_n| \geq N-2$, it follows trivially that $h = p$. □

Proposition 3.5.7. *Let s be a ϑ-standard word with seed and h be its critical integer. Any prefix p of s of length $> |B_h|$ has a ϑ-palindromic suffix with a unique occurrence in p.*

Proof. Since $|p| > |B_h|$ there exists $n \geq h$ such that

$$|B_n x_n| \leq |p| < |B_{n+1}| \ ,$$

with $B_{n+1} = (B_n x_n)^\oplus$ by the definition of h.

We can write $B_n x_n = vQ$, where Q is the longest ϑ-palindromic suffix of $B_n x_n$, which is nonempty, and, as shown in the proof of Lemma 3.5.3, has a unique occurrence in $B_n x_n$. Since $B_{n+1} = vQ\bar{v}$, we can write $p = vQ\bar{v}_2$, where $v = v_1 v_2$ for some $v_1, v_2 \in A^*$ and $|v_2| < |v|$. Now $v_2 Q \bar{v}_2$ is a ϑ-palindromic suffix of p which has a unique occurrence in p, for otherwise Q would be repeated in $B_n x_n$. This concludes the proof. □

Let us observe that in the case of a standard Episturmian word s, a stronger result holds: any prefix p of s has a palindromic suffix which is unrepeated in p (cf. [29]).

Proposition 3.5.8. *Let s be a ϑ-standard word with seed, and h be its critical integer. For any ϑ-palindromic factor P of length $|P| > |B_h|$, every first return to P in s is a ϑ-palindrome.*

Proof. Let P be a ϑ-palindromic factor of s, with $|P| > |B_h|$. Let $u \in \mathrm{Fact}(s)$ be a first return to P, i.e., $u = P\lambda = \rho P$, $\lambda, \rho \in A^*$, and the only two occurrences of P in u are as a prefix and as a suffix of u. If $|P| > |\rho|$, then the prefix P of u overlaps with the suffix P in u and this implies, as is easily to verify, that u is a ϑ-palindrome. Then let us suppose that $u = PvP$ with $v \in A^*$.

Now we consider the first occurrence of u or of $\bar{u}$ in s. Without loss of generality, we may suppose that $s = \alpha u s'$, and $\bar{u}$ does not occur in the prefix of s having length $|\alpha u| - 1$. Let Q be the ϑ-palindromic suffix of αu of maximal length. If $|Q| > |u|$, then we have that $\bar{u}$ occurs in αu before u, which is absurd. Then suppose $|Q| \leq |u|$. If $|u| > |Q| > |P|$, then one contradicts the hypothesis that u is a first return to P. If $|Q| = |P|$, then $Q = P$ has more than one occurrence in αu. Since $|\alpha u| > |B_h|$, one reaches a contradiction by Proposition 3.5.7. Thus the only remaining possibility is $Q = u$, i.e., u is a ϑ-palindrome. $\square$

In the case of Episturmian words, one has the stronger result that *every* first return to a palindrome is a palindrome. This was proven in [3] (see also [13]). However this cannot be extended to ϑ-Episturmian words. For instance, let s be the standard ϑ-Episturmian word $(abaca)^\omega$, where $\vartheta(a) = a$ and $\vartheta(b) = c$. Then aba is a first return to a in s, but it is not a ϑ-palindrome.

3.6 Structure of ϑ-Episturmian words

In this section we shall analyse in detail the class of ϑ-Episturmian words, also by showing some relations with the other classes introduced so far.

From Corollary 3.5.6 and Theorem 3.3.15, one derives the following

Proposition 3.6.1. *Let s be a standard ϑ-Episturmian word, h be its critical integer, and $x = x_1x_2\cdots x_n\cdots$ be the subdirective word of s. Then s is the image, by an injective morphism, of the standard Episturmian word t whose directive word is $x_{h+1}x_{h+2}\cdots x_n\cdots$.*

However, this can be improved. In fact, the next results will show (cf. Theorem 3.4.11) that every $s \in SEpi_\vartheta$ is a morphic image, by an injective morphism, of the standard Episturmian word whose directive word is precisely x, the subdirective word of s.

In the following we shall denote by $\mathcal{P}_\vartheta$, or simply $\mathcal{P}$, the set of unbordered ϑ-palindromes. We remark that $\mathcal{P}$ is a *biprefix code*, i.e., none of its elements is a proper prefix or suffix of other elements of $\mathcal{P}$.

Proposition 3.6.2. *$PAL_\vartheta^* = \mathcal{P}^*$.*

Proof. Since $\mathcal{P} \subseteq PAL_\vartheta$, one has $\mathcal{P}^* \subseteq PAL_\vartheta^*$. Thus it suffices to show that every nonempty ϑ-palindrome admits a factorization in unbordered ϑ-palindromes, i.e., is in $\mathcal{P}^*$. Note that such a factorization is necessarily unique, as $\mathcal{P}$ is a code.

Let $w \in PAL_\vartheta$. If $|w| = 1$, then clearly w is unbordered, so that $w \in \mathcal{P}$. Let then $|w| > 1$ and suppose, by induction, that every ϑ-palindrome which is shorter than w can be factorized in elements of $\mathcal{P}$. If w is unbordered, then we are done. Let then u be the longest proper border of w. Since w is a ϑ-palindrome, so is u.

If $|w| \geq 2|u|$, then $w = uvu$ for some $v \in PAL_\vartheta$, so that both $u, v \in \mathcal{P}^*$ by induction. This implies the assertion in this case.

If $|w| < 2|u|$, then there exists a border β of u such that $w = u_1\beta\bar{u}_1$, where $u = u_1\beta = \beta\bar{u}_1$. By induction, both β and $u = u_1\beta$ are in $\mathcal{P}^*$; since $\mathcal{P}$ is a biprefix code, this implies that $u_1 = u\beta^{-1}$ is in $\mathcal{P}^*$ too. Hence $w = u_1 u \in \mathcal{P}^*$ as requested. ☐

Example 3.6.3. Let $A = \{a, b, c, d, e\}$ and ϑ be the antimorphism defined by $\bar{a} = a$, $\bar{b} = c$, and $\bar{d} = e$. The word $acbdaaecba.abaca \in PAL_\vartheta^2$ can be uniquely factorized in unbordered ϑ-palindromes as:

$$a.cb.daae.cb.a.a.bac.a\ .$$

We remark that from the preceding proposition one derives that any standard ϑ-Episturmian word s admits a (unique) infinite factorization in elements of $\mathcal{P}$, i.e., one can write

$$s = \pi_1\pi_2\cdots\pi_n\cdots\ , \qquad \text{with } \pi_i \in \mathcal{P} \text{ for all } i > 0\ . \tag{3.19}$$

Lemma 3.6.4. *Let $s \in SEpi_\vartheta$, with $s = \pi_1\pi_2\cdots\pi_n\cdots$ as above. Let u be a nonempty and proper prefix of π_n, for some $n > 0$. Then u is not right special in s.*

Proof. By contradiction, assume that u is a right special factor of s. Then it is not left special; indeed, otherwise it would be a ϑ-palindrome since s is ϑ-Episturmian, and this is clearly absurd as $\pi_n \in \mathcal{P}$.

Consider now the smallest integer h such that u is a prefix of π_h. If $h = 1$, then u would be a ϑ-palindrome, which is again a contradiction. Let then $h > 1$. Since u is not left special, $\bar{a}_{h-1}u$ is its unique left extension in s. One can keep extending to the left in a unique way, until one gets a left special factor, or reaches the beginning of the word. In either case, the factor q of s that one obtains is a prefix of s. Moreover it is right special in s, as every occurrence of the right special factor u extends to the left to q. Hence $\bar{q}$ is a left special factor of s, and then a prefix of s. Thus q is a ϑ-palindrome, and therefore it begins with $\bar{u}$. One has $|q| \geq 2|u|$, for otherwise there would be a nonempty word in $\mathrm{Pref}(u) \cap \mathrm{Suff}(\bar{u})$, that is, a nonempty ϑ-palindromic prefix of u, which contradicts the hypothesis that u is a proper prefix of π_h. Thus $q = \bar{u}q'u$ for some $q' \in PAL_\vartheta$.

We have $\pi_1 \cdots \pi_{h-1} \in \mathcal{P}^*$ and, by Proposition 3.6.2, $q' \in \mathcal{P}^*$. Since $\mathcal{P}$ is a biprefix code, this implies $\pi_1 \cdots \pi_{h-1}(q')^{-1} \in \mathcal{P}^*$, i.e., $q' = \pi_{h'} \cdots \pi_{h-1}$ for some $h' \leq h$ (if $h' = h$, then $q' = \varepsilon$). Then $\pi_1 \cdots \pi_{h'-1}$ has $\bar{u}$ as a suffix. As $\bar{u}$ has no nonempty ϑ-palindromic suffixes, it is a proper suffix of $\pi_{h'-1}$, which then begins in u, contradicting the minimality of h. □

The next result has already been proved (cf. Theorem 3.4.11) however, we report here a different proof for the sake of completeness.

Theorem 3.6.5. *Let $s \in A^\omega$ be a standard ϑ-Episturmian word, Δ be its subdirective word, and $B = \mathrm{alph}(\Delta)$. There exists an injective morphism $\mu : B^* \to A^*$ such that $s = \mu(\psi(\Delta))$ and $\mu(B) \subseteq \mathcal{P}$.*

Proof. We can assume that s can be factorized as in (3.19). For any $n \geq 0$, let a_n be the first letter of π_n. We shall prove that if $n, m \geq 0$ are such that $a_n = a_m$, then $\pi_n = \pi_m$.

Let u be the longest common prefix of π_n and π_m, which is nonempty as $a_n = a_m$. By contradiction, suppose $\pi_n \neq \pi_m$. Then, as $\mathcal{P}$ is a biprefix code, u must be a *proper* prefix of both π_n and π_m, so that there exist two distinct letters b_n, b_m such that ub_n is a prefix of π_n and ub_m is a prefix of π_m. Hence u is a right special factor of s, but this contradicts the previous lemma.

We have shown that for any $n > 0$, π_n is determined by its first letter a_n. Thus, letting

$$C = \{a_n \mid n > 0\} \subseteq A ,$$

it makes sense to define an injective morphism $\mu : C^* \to A^*$ by setting $\mu(a_n) = \pi_n$ for all $n > 0$. The word

$$t = \mu^{-1}(s) = a_1 a_2 \cdots a_n \cdots \in C^\omega$$

has infinitely many palindrome prefixes, each being the inverse image of a ϑ-palindromic prefix of s. Indeed, if $\pi_1 \cdots \pi_n$ is a ϑ-palindromic prefix of s, by the uniqueness of the factorization over $\mathcal{P}$ one obtains $\pi_i = \pi_{n+1-i}$ for $i = 1, \ldots, n$; conversely, if $w \in PAL$, then trivially $\mu(w) \in PAL_\vartheta$. Hence t is closed under reversal.

Let w be a left special factor of t, and let i, j be such that $a_i \neq a_j$ and $a_i w, a_j w \in \mathrm{Fact}(t)$. Then $\bar{a}_i \mu(w), \bar{a}_j \mu(w) \in \mathrm{Fact}(s)$, so that $\mu(w)$ is a left special factor of s, and hence a prefix of it. Again by the uniqueness of the factorization of s over the prefix code $\mathcal{P}$, one derives $w \in \mathrm{Pref}(t)$. Therefore t is a standard Episturmian word over C.

Let $\Delta = x_1 x_2 \cdots x_n \cdots$, and let $B_n = \mu(a_1) \cdots \mu(a_{r_n})$ be the n-th ϑ-palindromic prefix of s for any $n > 1$. Then, as shown above, $a_1 \cdots a_{r_n}$ is exactly the n-th palindromic prefix of t. Since the only word occurring in the factorization (3.19) and beginning with x_n is $\mu(x_n)$, we have $B_n \mu(x_n) \in \mathrm{Pref}(s)$, so that $x_n = a_{r_n+1}$ for all $n > 1$. This proves that the directive word of t is exactly Δ, and hence $C = B$. $\square$

Corollary 3.6.6. *A standard ϑ-Episturmian word s is ϑ-standard if and only if $s = \mu_\vartheta(t)$ for some $t \in A^\omega$.*

Proof. If s is ϑ-standard, then by Theorem 3.2.2 there exists a standard Episturmian word t such that $s = \mu_\vartheta(t)$. Conversely, if $t \in A^\omega$ and $s = \mu_\vartheta(t)$, then, since $\mu_\vartheta(a) \in \mathcal{P}$ for any $a \in A$, by the uniqueness of the factorization over $\mathcal{P}$ one has that μ_ϑ is the morphism μ considered in the preceding theorem. Thus $t = \mu_\vartheta^{-1}(s)$ is a standard Episturmian word and s is ϑ-standard by Theorem 3.2.2. $\square$

Proposition 3.6.7. *Let $\mu : B^* \to A^*$ be a nonerasing morphism such that*

1. $\mu(x) \in PAL_\vartheta$ *for all* $x \in B$*,*

2. $\mathrm{alph}(\mu(x)) \cap \mathrm{alph}(\mu(y)) = \emptyset$ *if* $x, y \in B$ *and* $x \neq y$*,*

3. $|\mu(x)|_a \leq 1$ *for all* $x \in B$ *and* $a \in A$*.*

Then, for any standard Episturmian word $t \in B^\omega$*,* $s = \mu(t)$ *is a standard* ϑ*-Episturmian word.*

Proof. From the first condition one obtains that μ sends palindromes into ϑ-palindromes, so that s has infinitely many ϑ-palindromic prefixes, and is therefore closed under ϑ.

Let w be a nonempty left special factor of s. Suppose first that w is a proper factor of $\mu(x)$ for some $x \in B$, and is not a prefix of $\mu(x)$. Let a be the first letter of w. By the second condition, $\mu(x)$ is the only word in $\mu(B)$ containing the letter a; by condition 3, a occurs exactly once in $\mu(x)$. Since a is not a prefix of $\mu(x)$, it is always preceded in s by the letter which precedes a in $\mu(x)$. Hence a is not left special, a contradiction.

Thus we can write w as $w_1\mu(u)w_2$, where w_1 is a proper suffix of $\mu(x_1)$ and w_2 is a proper prefix of $\mu(x_2)$, for some suitable $x_1, x_2 \in B$ such that $x_1ux_2 \in \mathrm{Fact}(t)$. One can prove that $w_1 = \varepsilon$ by showing, as done above, that otherwise its first letter, which would not be a prefix of $\mu(x_1)$, could not be left special in s.

Therefore $w = \mu(u)w_2$. Reasoning as above, one can prove that if $w_2 \neq \varepsilon$, then w is not right special, and more precisely that each occurrence of w can be extended on the right to an occurrence of $\mu(ux_2)$. Since w is left special in s, so is $\mu(ux_2)$.

Without loss of generality, we can then suppose $w = \mu(u)$. Since μ is injective by condition 2, u is uniquely determined. As w is left special in s, there exist two letters $a, b \in A$, $a \neq b$, such that $aw, bw \in \mathrm{Fact}(s)$. Hence there exist two (distinct) letters $x_a, x_b \in B$ such that $x_au, x_bu \in \mathrm{Fact}(t)$. Then u is a left special factor of t and hence a prefix of t, so that $w = \mu(u)$ is a prefix of s. □

Example 3.6.8. Consider the standard Sturmian word

$$t = aabaaabaaabaab\cdots$$

having the directive word $(aab)^\omega$. Let $A = \{a, b, c, d, e\}$, and ϑ be the involutory antimorphism defined by $\bar{a} = b$, $\bar{c} = c$, $\bar{d} = e$. If μ is the morphism $\mu : \{a, b\}^* \to A^*$ defined by $\mu(a) = acb$ and $\mu(b) = de$, then the word

$$s = \mu(t) = acbacbdeacbacbacbde \cdots$$

is a standard ϑ-Episturmian word. We observe that s is not ϑ-standard, since it does not begin with $ab = a^\oplus$.

Remark. Any morphism satisfying conditions 1 and 3 in the statement of Proposition 3.6.7 is such that $\mu(x) \in \mathcal{P}$ for any letter x. However there exist standard ϑ-Episturmian words for which the morphism μ given by Theorem 3.4.11 does not satisfy condition 3. For instance, the standard ϑ-Episturmian word $s = (abaca)^\omega$, with $\bar{a} = a$ and $\bar{b} = c$, is given by $s = \mu(t)$, where $t = \psi(aba^\omega)$, $\mu(a) = a$, and $\mu(b) = bac$.

We say that a subset B of the alphabet A is *ϑ-skew* if $B \cap \vartheta(B) \subseteq PAL_\vartheta$, that is, if

$$x \in B,\ x \neq \bar{x} \Longrightarrow \bar{x} \notin B\ . \tag{3.20}$$

Proposition 3.6.9. *Let s be a standard ϑ-Episturmian word and Δ be its subdirective word. Then $B = \mathrm{alph}(\Delta)$ is ϑ-skew.*

Proof. We can factorize s as in (3.19). By Theorem 3.4.11, it suffices to show that if $\pi_n = xw\bar{x}$ for some $n > 0$ and $w \in A^*$, then π_k does not begin with $\bar{x}$, for any $k > 0$. By contradiction, let k be the smallest integer such that $\bar{x} \in \mathrm{Pref}(\pi_k)$. Without loss of generality, we can assume $n < k$. By Lemma 3.6.4, no suffix of $w\bar{x}$ is a left special factor of s. Hence every occurrence of $\bar{x}$ in s is preceded by xw (or by a proper suffix of it, if the beginning of the word is reached). First suppose that π_k is preceded in s by xw. Then, since $w \in PAL_\vartheta \subseteq \mathcal{P}^*$ and $\mathcal{P}$ is a biprefix code, one has $w = \pi_{k'} \cdots \pi_{k-1}$ for some $k' \leq k$. Thus $\pi_{k'-1}$ ends in x and therefore begins with $\bar{x}$, contradicting the minimality of k.

If $\pi_1 \cdots \pi_{k-1} \in \mathrm{Suff}(w)$, from $n < k$ it follows that $\pi_n = xw\bar{x}$ is a proper factor of itself, which is trivially absurd. $\square$

A ϑ-standard word s can have left special factors which are not prefixes of s. Such factors have length at most 2, by Theorem 3.2.7. For instance, consider the ϑ-standard word s with $\vartheta = E \circ R$ and $\Delta(s) = (ab)^\omega$. One has $s = abbaababbaabbaab\cdots$. As one easily verifies, b and ba are two left special factors which are not prefixes. Hence in general, a ϑ-standard word is not standard ϑ-Episturmian.

The next proposition gives a characterization of ϑ-standard words which are standard ϑ-Episturmian.

Proposition 3.6.10. *A ϑ-standard word s is standard ϑ-Episturmian if and only if $B = \mathrm{alph}(\Delta(s))$ is ϑ-skew.*

Proof. Let s be a ϑ-standard word such that B is ϑ-skew. By Theorem 3.2.2, one has $s = \mu_\vartheta(t)$, where $t = \psi(\Delta(s))$ is a standard Episturmian word. The morphism μ_ϑ satisfies conditions 1 and 3 in Proposition 3.6.7 by definition. By (3.20), one easily derives that the restriction of μ_ϑ to $\mathrm{alph}(t) = B$ satisfies also condition 2 of Proposition 3.6.7, so that $s = \mu_\vartheta(t)$ is a standard ϑ-Episturmian word.

The converse is a consequence of Proposition 3.6.9, as the subdirective word of a ϑ-standard word s is $\Delta(s)$. $\square$

Example 3.6.11. Let $A = \{a, b, c, d, e\}$, $\Delta = (acd)^\omega$, and ϑ be defined by $\bar{a} = b$, $\bar{c} = c$, and $\bar{d} = e$. The ϑ-standard word $\psi_\vartheta(\Delta) = abcabdeabcaba\cdots$ is standard ϑ-Episturmian.

Let us observe that in general a standard ϑ-Episturmian word is not a ϑ-standard word. A simple example is given by the word $s = (abaca)^\omega$, where ϑ is the antimorphism which exchanges b with c and fixes a. One easily verifies that ε and a are the only left special factors of s, so that s is standard ϑ-Episturmian. However (cf. Proposition 3.2.3) s is not ϑ-standard, since ab is a prefix of s, but $(ab)^\oplus = abca$ is not. Another example is the word s considered in Example 3.6.8: s is standard ϑ-Episturmian, but it is not ϑ-standard because its first nonempty ϑ-palindromic prefix is acb and not $ab = a^\oplus$.

Although neither of the two classes (ϑ-standard and standard ϑ-episturmian words) is included in the other one, the following relation holds.

Proposition 3.6.12. *Every ϑ-standard word is a morphic image, under a literal morphism, of a standard $\hat{\vartheta}$-Episturmian word, where $\hat{\vartheta}$ is an extension of ϑ to a larger alphabet.*

Proof. Let $s = \psi_\vartheta(\Delta)$ be a ϑ-standard word, $B \subseteq A$ be the set of letters occurring in Δ, and $A' = A \setminus PAL_\vartheta$. Moreover, let us set

$$C = \{c \in B \cap A' \mid \exists r \in (B \setminus \{c, \bar{c}\})^* : r\bar{c} \in \mathrm{Pref}(\Delta)\} ,$$

i.e., C is the set of letters c occurring in Δ and such that $\bar{c}$ occurs before the first occurrence of c. If $C = \emptyset$, then by the previous proposition s is a standard ϑ-Episturmian word, so that the assertion is trivially verified. Let us explicitly note that if $c \in C$, then $\bar{c} \notin C$.

Suppose then C nonempty, and let $C' = \{c' \mid c \in C\}$ and $\hat{C} = \{\hat{c} \mid c \in C\}$ be two sets having the same cardinality as C, both disjoint from A. One can then naturally define the bijective map $\varphi : B \to (B \setminus C) \cup C'$ such that $\varphi(a) = a$ if $a \notin C$, and $\varphi(a) = a'$ otherwise. Set $\hat{A} = A \cup C' \cup \hat{C}$, and define an involutory antimorphism $\hat{\vartheta}$ over $\hat{A}$ by setting $\hat{\vartheta}_{|A} = \vartheta$ and $\hat{\vartheta}(c') = \hat{c}$ for any $c' \in C'$.

Extending φ to a morphism from B^* to $\hat{A}^*$, it makes sense to consider the infinite word $\hat{\Delta} = \varphi(\Delta)$ over $\hat{A}$. Thus we can define as well the $\hat{\vartheta}$-standard word $\hat{s}$ directed by $\hat{\Delta}$. Since $\mathrm{alph}(\hat{\Delta})$ is $\hat{\vartheta}$-skew, by the previous proposition $\hat{s}$ is also $\hat{\vartheta}$-Episturmian.

By Theorem 3.2.2, one has $s = \mu_\vartheta(\psi(\Delta))$ and $\hat{s} = \mu_{\hat{\vartheta}}(\psi(\hat{\Delta}))$. Since φ is injective on B, it follows $\psi(\hat{\Delta}) = \varphi(\psi(\Delta))$, so that

$$\hat{s} = \mu_{\hat{\vartheta}}(\varphi(\psi(\Delta))) . \tag{3.21}$$

Let $g : \hat{A}^* \to A^*$ be the literal morphism defined as follows:

$$g_{|C'} = \varphi^{-1}, \quad g_{|\hat{C}} = \vartheta \circ \varphi^{-1} \circ \hat{\vartheta}, \ \text{ and } g_{|A} = \mathrm{id} ,$$

i.e., $g(a) = a$ if $a \in A$, and for all $c \in C$, $g(c') = c$ and $g(\hat{c}) = \bar{c}$. We want to show that $g(\hat{s}) = s = \mu_\vartheta(\psi(\Delta))$. In view of (3.21), it suffices to prove that $g \circ \mu_{\hat{\vartheta}} \circ \varphi = \mu_\vartheta$ over B. Indeed, by the definitions, if $c \in C$ then

$$g(\mu_{\hat{\vartheta}}(\varphi(c))) = g(c'\hat{c}) = c\bar{c} = \mu_\vartheta(c) ,$$

whereas if $a \in B \setminus C$, then

$$g(\mu_{\hat{\vartheta}}(\varphi(a))) = g(a^\oplus) = a^\oplus = \mu_\vartheta(a) . \qquad \square$$

Example 3.6.13. Let $A = \{a, b\}$, $\vartheta = E \circ R$ (i.e., $\bar{a} = b$), and s be the ϑ-standard word having the directive sequence $\Delta = (ab)^\omega$, so that

$$s = abbaababbaabbaab\cdots .$$

In this case $A' = A = B$, $C = \{b\}$, $C' = \{b'\}$, and $\hat{C} = \{\hat{b}\}$. We set $c = b'$ and $d = \hat{b}$, so that $\hat{A} = \{a, b, c, d\}$, $\hat{\vartheta}(a) = b$, and $\hat{\vartheta}(c) = d$. The morphism φ in this case is defined by $\varphi(a) = a$ and $\varphi(b) = c$. Hence $\hat{\Delta} = \varphi(\Delta) = (ac)^\omega$. The $\hat{\vartheta}$-standard (and standard $\hat{\vartheta}$-Episturmian) word $\hat{s}$ directed by $\hat{\Delta}$ is

$$\hat{s} = abcdababcdabcdabab\cdots .$$

The literal morphism g is defined by $g(a) = g(d) = a$, and $g(b) = g(c) = b$. One has $g(\hat{s}) = s$.

3.7 The importance of involutory antimorphisms

The main result of this section shows that the existence of an underlying involutory antimorphism θ is a consequence of three natural word combinatorial assumptions: recurrence, uniqueness of right and left special factors, and constant growth of the factor complexity:

Theorem 3.7.1. *Let $\omega \in A^{\mathbb{N}}$ be a word on a finite alphabet A. Suppose*

1. *ω is recurrent.*
2. *For each $n \geq 1$, ω has a unique right special factor of length n and a unique left special factor of length n.*
3. *There exists a constant K such that $p(n) = \mathrm{card} A + (n-1)K$ for each $n \geq 1$.*

Then there exists an involutory antimorphism $\theta : A^* \to A^*$ relative to which ω is a θ-Episturmian word.

While each of the hypotheses (1)–(3) above is in fact necessary (see the examples below), Theorem 3.7.1 is not a characterization of θ-Episturmian words since the converse is in general false. For instance, it is easy to verify that the word on $\{a, b, c\}$ obtained by applying the morphism $0 \mapsto a$ and $1 \mapsto bac$ to the Fibonacci word $\mathbf{f}$ does not satisfy condition (3) above but is θ-Episturmian relatively to the involutory antimorphism generated by $\theta(a) = a$ and $\theta(b) = c$.

The next series of examples illustrate that each of the hypotheses (1)–(3) above is in fact necessary and independent of one another. In what follows $\mathbf{f}$ denotes the Fibonacci infinite word.

Example 3.7.2. The word $2\mathbf{f} = 20100101001001010010100100 1010\cdots$ satisfies conditions (2) and (3) but not (1). The set of factors of this word is not closed under θ for any choice of the involutory antimorphism θ of $\{0, 1, 2\}$, so that $2\mathbf{f}$ is not θ-Episturmian.

Example 3.7.3. The fixed point of the morphism $0 \mapsto 021$, $1 \mapsto 0$, $2 \mapsto 01$ satisfies (1) and (3) but not (2), in fact for each $n \geq 1$, this word has a unique right special factor of length n but two distinct left special factors of length n. Hence this word is not θ-Episturmian.

Example 3.7.4. Consider the word $\omega = \tau \circ \sigma(\mathbf{f})$ where $\sigma(0) = 0$, $\sigma(1) = 12$, $\tau(0) = 10$, $\tau(1) = 1$, and $\tau(2) = 12$. It is readily verified that ω satisfies conditions (1) and (2), but not (3) as $p(1) = 3$, $p(2) = 5$, and $p(3) = 6$. The word ω is not θ-Episturmian, in fact one easily verifies that the factor 10112101 is a bispecial factor of ω and yet is not fixed by any involutory antimorphism.

Using the notion of degree, condition (3) in Theorem 3.7.1 can be replaced by the following: *All nonempty right special factors and all nonempty left special factors of ω have the same degree, namely $K + 1$* (cf. Lemma 3.7.5 in next section). We remark that in the case $K = \mathrm{card}A - 1$ condition (3) is trivially true also for $n = 0$, and conditions (1)–(3) give a characterization of Arnoux-Rauzy words.

For definitions and notations not given in the text the reader is referred to [44, 8, 13, 12].

3.7.1 Proof of Theorem 3.7.1

The proof is organized as follows. First we prove that any factor of ω is contained in a bispecial factor of ω. In particular, this implies that ω has infinitely many distinct bispecial factors. Next, we prove that there exists an involutory antimorphism θ of A^* such that all bispecial factors are θ-palindromes. From this we derive that θ preserves the set of factors of ω, so that ω is ϑ-Episturmian.

The following notation will be useful in the proof of Theorem 3.7.1: Let u and v be non-empty factors of ω. We write $u \vdash uv$ to mean that for each factor w of ω with $|w| = |u| + |v|$, if w begins in u then $w = uv$. If it is not the case that $u \vdash uv$, then we will write $u \nvdash uv$. Similarly we will write $vu \dashv u$ to mean that for each factor of ω with $|w| = |u| + |v|$ if w ends in u then $w = vu$. Otherwise we write $vu \not\dashv u$.
We begin with a few lemmas. The following lemma is an immediate consequence of the hypotheses of Theorem 3.7.1:

Lemma 3.7.5. *Let u and u' be right (respectively left) special factors of ω. Then under the hypotheses of Theorem 3.7.1, for any letter $a \in A$, ua (respectively au) is a factor of ω if and only if $u'a$ (respectively au') is a factor of ω.*

Proof. Conditions (2) and (3) of Theorem 3.7.1 imply that K is a positive integer, and that each right special factor u has exactly $K + 1$ distinct right extensions of the form ua with $a \in A$, i.e., has degree $K + 1$. Moreover, if u and u' are right special factors of ω, then by (2) one is a suffix of the other. Hence ua is a factor of ω if and only if $u'a$ is a factor of ω. A similar argument applies to left special factors of ω. □ □

Lemma 3.7.6. *Let u be a factor of ω. Then under the hypotheses of Theorem 3.7.1 we have that u is a factor of a bispecial factor of ω. Let W denote the shortest bispecial factor of ω containing u. Then u occurs exactly once in W.*

Proof. We first observe that by condition (2) of Theorem 3.7.1, ω is not periodic.

Since ω is recurrent, there exists a factor z of ω which begins and ends in u and has exactly two occurrences of u. Writing $z = vu$, clearly we have $vu \not\dashv u$, otherwise ω would be periodic. Thus some suffix of z of length at least $|u|$ must be a left special factor of ω. Let $x \in A^*$ be of minimal length such that xu is a left special factor of ω. Such a word is trivially unique, and we have

$xu \dashv u$. In a similar way, there exists a unique $y \in A^*$ of minimal length such that uy is right special in ω, and it satisfies $u \vdash uy$.

From the preceding relations one obtains $xu \vdash xuy$ and $xuy \dashv uy$. Since xu is left special in ω and xu is always followed by y one has that xuy is also left special. Similarly, since uy is right special and always preceded by x, xuy is right special. Hence every factor u of ω is contained in some bispecial factor $W = xuy$ of ω. Furthermore, this W is the shortest bispecial factor containing u. Indeed, if $W' = x'uy'$ is bispecial in ω and $|W'| < |W|$, then either $|x'| < |x|$ or $|y'| < |y|$; since $x'u$ and uy' are respectively a left and a right special factor of ω, this violates the minimality of x or y. Using the same argument, one shows that W cannot have more than one occurrence of u. □ □

It follows immediately from Lemma 3.7.6 that ω, under the hypotheses of Theorem 3.7.1, contains an infinite number of distinct bispecial factors

$$\varepsilon = W_0, W_1, W_2, \ldots$$

which we write in order of increasing length. Thus, as a consequence of condition (2), for each $k \geq 1$ we have that W_{k+1} begins and ends in W_k.

Lemma 3.7.7. *Let $a \in A$, and let W_k be the shortest bispecial factor of ω containing a. Then $W_k = W_{k-1}VW_{k-1}$, where V contains the letter a. Moreover, all letters in V are distinct and none of them occurs in W_{k-1}. If Ua is a factor of ω for some bispecial factor U, then a is the first letter of V.*

Proof. Clearly since W_k begins and ends in W_{k-1} and a does not occur in W_{k-1}, it follows that $W_k = W_{k-1}VW_{k-1}$, for some non-empty factor V containing a. We will first show that the first letter of V does not occur in W_{k-1}. Then we will show that no letter of V occurs in W_{k-1}. Thus for each letter b which occurs in V, we have that W_k is the shortest bispecial factor containing b. Hence by Lemma 3.7.6 we have that b occurs exactly once in V.

Let a' denote the first letter of V which does not occur in W_{k-1}. We claim that a' is the first letter of V. The result is clear in case $W_{k-1} = \varepsilon$. Thus we can assume W_{k-1} is non-empty. Suppose to the contrary that a' is not the first letter of V. Then there exists a letter b immediately preceding a' in V, which also occurs in W_{k-1}. We claim b is a right special factor of ω. This is trivial if

b is the last letter of W_{k-1}. If this is not true, then there is an occurrence of b in W_{k-1} followed by some letter $c \neq a'$. Thus b is a right special factor of ω.

Now, since ba' is a factor of ω, it follows from Lemma 3.7.5 that $W_k a'$ is a factor of ω. We can write $W_k a' = W_{k-1} X a' Y W_{k-1} a'$, with X non-empty. By the definition of a', one has that W_k is the shortest bispecial factor of ω containing a'. It follows that every occurrence of a' in ω is preceded by $W_{k-1}X$. Hence $W_{k-1}X$ is both a prefix and a suffix of W_k, whence is a bispecial factor of ω of length greater than $|W_{k-1}|$ and less than $|W_k|$, a contradiction. Hence a' is the first letter of V, in other words the first letter of V does not occur in W_{k-1}.

We next show that no letter in V occurs in W_{k-1}. Again this is clear in case $W_{k-1} = \varepsilon$. Thus we can assume W_{k-1} is non-empty. Suppose to the contrary: Let d denote the first letter in V which also occurs in W_{k-1}. We saw earlier that d is not the first letter of V. Thus the letter e preceding d in V does not occur in W_{k-1}. We claim that d is a left special factor, or equivalently is the first letter of W_{k-1}. Otherwise, if d were not the first letter of W_{k-1}, there would be an occurrence of d in W_{k-1} preceded by some letter $e' \neq e$. Thus d is left special, a contradiction.

Since ed is a factor of ω, it follows from Lemma 3.7.5 that eW_k is a factor of ω. We can write $eW_k = eW_{k-1}X'eY'W_{k-1}$ with Y' non-empty (since it contains d). Since e does not occur in W_{k-1}, it follows that W_k is the shortest bispecial factor of ω containing e, and hence every occurrence of e in ω is followed by $Y'W_{k-1}$. Hence $Y'W_{k-1}$ is both a prefix and a suffix of W_k, and hence a bispecial factor of ω whose length is greater than that of W_{k-1} but smaller than that of W_k. A contradiction. Hence, no letter occurring in V occurs in W_{k-1}.

Finally suppose Ua is a factor of ω for some bispecial factor U. By Lemma 3.7.5 we have that $W_k a$ is a factor of ω. Writing $W_k a = W_{k-1}X''aY''W_{k-1}a$, we have that every occurrence of a in ω is preceded by $W_{k-1}X''$, whence $W_{k-1}X''$ is both a prefix and a suffix of W_k. This implies that $W_{k-1}X''$ is a bispecial factor of ω, and hence equal to W_{k-1}. Thus X'' is empty and a is the first letter of V as required. This concludes the proof of Lemma 3.7.7. □ □

We now proceed with the proof of Theorem 3.7.1. It suffices to show that there exists an involutory antimorphism $\theta : A^* \to A^*$ relative to which each W_k is a θ-palindrome. Indeed, by Lemma 3.7.6 any factor u of ω is contained

in some W_k, and hence so is $\theta(u)$.
We proceed by induction on k. By Lemma 3.7.6, W_1 is of the form $W_1 = a_0a_1\cdots a_n$ with $a_i \in A$, $0 \leq i \leq n$, and with $a_i \neq a_j$ for $i \neq j$. Hence we can begin by defining θ on the subset $\{a_0, a_1, \ldots, a_n\}$ of A, by $\theta(a_i) = a_{n-i}$. Thus $\theta(W_1) = W_1$, i.e., W_1 is a θ-palindrome.

By induction hypothesis, let us assume θ is defined on the set of all letters occurring in $W_1, W_2, \ldots, W_k$ with each W_i $(1 \leq i \leq k)$ a θ-palindrome. Let $a \in A$ be the unique letter such that W_ka is a prefix of W_{k+1} and then a left special factor of ω. We consider two cases: Case 1: a does not occur in W_k, and Case 2: a occurs in W_k.

Case 1: Since a, does not occur in W_k but occurs in W_{k+1}, it follows from Lemma 3.7.7 that $W_{k+1} = W_kVW_k$ where all letters of V are distinct and none of them occurs in W_k. Thus we can write $V = b_0b_1\cdots b_{|V|-1}$ and extend the domain of definition of θ to $\{b_0, b_1, \ldots, b_{|V|-1}\}$ by $\theta(b_i) = b_{|V|-i-1}$. In this way W_{k+1} becomes a θ-palindrome.

Case 2: In this case we will show that W_{k+1} is the *θ-palindromic closure* of W_ka, that is the shortest θ-palindrome beginning in W_ka. In fact we will show that $W_{k+1} = W_kaV$ where $W_k = UaV$ for some word V and θ-palindrome U.

Let W_n be the shortest bispecial factor containing a. Hence $n \leq k$. Since W_ka is a factor of ω, it follows from Lemma 3.7.7 that $W_{n-1}a$ is a prefix of W_n, and hence a prefix of W_k. Thus there exists a bispecial factor U (possibly empty) such that Ua is a prefix of W_k. Let U denote the longest bispecial factor of ω with the property that Ua is a prefix of W_k, and write $W_k = UaV$, where V is possibly the empty word. We will show that $W_{k+1} = W_kaV$.

Setting $\bar{a} = \theta(a)$, we will show that $\bar{a}Ua \vdash \bar{a}UaV$. First of all, since Ua is a prefix of the θ-palindrome W_k, and U is bispecial and then θ-palindrome, it follows that $\bar{a}U$ is a factor of ω; hence by Lemma 3.7.5, $\bar{a}W_k = \bar{a}UaV$ is a factor of ω. Suppose to the contrary that $\bar{a}Ua \nvdash \bar{a}UaV$. Then there exists a proper prefix V' of V and a letter $b \in A$ such that $V'b$ is not a prefix of V and $\bar{a}UaV'b$ is a factor of ω. Thus $\bar{a}UaV'$ is right special, and hence $\bar{a}UaV'$ is a suffix of W_k. Since UaV' is also a prefix of W_k, it follows that UaV' is bispecial, and hence a θ-palindrome. We deduce that $UaV'a$ is a prefix of W_k contradicting the maximality of the length of U. Thus, $\bar{a}Ua \vdash \bar{a}UaV$ as required. It follows that $W_ka \vdash W_kaV$, since $\bar{a}Ua$ is a suffix of W_ka. Hence W_kaV is a left special factor of ω, as the $W_k a$ is left special and extends uniquely to $W_k aV$.

It remains to show that $W_k aV$ is also right special. In the same way that we showed that $\bar{a}Ua \vdash \bar{a}UaV$, a symmetric argument shows that $\theta(V)\bar{a}Ua \dashv \bar{a}Ua$. Thus to show that $W_k aV$ is right special, it suffices to show that $\bar{a}UaV$ is right special. Now since $W_k a$ is left special and $\bar{a}U$ is a factor of ω, it follows from Lemma 3.7.5 that $\bar{a}W_k a = \bar{a}UaVa$ is a factor of ω. So if $\bar{a}UaV$ were not right special, it would mean that $\bar{a}Ua \vdash \bar{a}UaV \vdash \bar{a}UaVa = \bar{a}\theta(V)\bar{a}Ua$. This implies that ω is periodic, a contradiction. Thus $W_k aV$ is right special, and hence bispecial. Since $W_k a \vdash W_k aV$, W_{k+1} cannot be a proper prefix of $W_k aV$, so that $W_{k+1} = W_k aV$.

It remains to show that W_{k+1} is a θ-palindrome. But, using the fact that U is a θ-palindrome, $\theta(W_{k+1}) = \theta(W_k aV) = \theta(V)\bar{a}W_k = \theta(V)\bar{a}UaV = \theta(V)\bar{a}\theta(U)aV = \theta(UaV)aV = \theta(W_k)aV = W_k aV = W_{k+1}$. Thus W_{k+1} is a θ-palindrome.

Having established that each bispecial factor of ω is a θ-palindrome, we conclude that ω is a θ-Episturmian word. This concludes the proof of Theorem 3.7.1.

Remark. It follows that for each $k \geq 1$, the θ-palindromic prefixes of W_k are precisely the bispecial prefixes of W_k.

Let θ be an involutory antimorphism of the free monoid A^*. In [12] the authors introduced various sets of words whose factors are closed under the action of θ. One such set is $SW_\theta(N)$ consisting of all infinite words ω whose sets of factors are closed under θ and such that every left special factor of ω of length greater or equal to N is a prefix of ω. Thus $SW_\theta(0)$ is precisely the set of all standard θ-Episturmian words. Fix $N \geq 0$, and let $\omega \in SW_\theta(N)$. Let $(W_n)_{n\geq 0}$ denote the sequence of all θ-palindromic prefixes of ω ordered by increasing length. For each $n \geq 0$ let $x_n \in A$ be such that $W_n x_n$ is a prefix of ω. The sequence $(x_n)_{n\geq 0}$ is called the *subdirective word* of ω. In [12], the authors establish the following lemma (Lemma 4.3 in [12]):

Lemma 3.7.8. *Let $\omega \in SW_\theta(N)$. Suppose $x_n = x_m$ for some $0 \leq m < n$ and with $|W_m| \geq N - 2$. Then $W_{n+1} = (W_n x_n)^{\oplus_\theta}$.*

In case $N = 0$, we can say more:

Proposition 3.7.9. *Let ω be a standard θ-Episturmian word. Suppose that $W_n a$ is left special for some $n > 0$, and that the letter a occurs in W_n. Then $W_{n+1} = (W_n a)^{\oplus_\theta}$.*

Proof. By Lemma 3.7.8 it suffices to show that for some $0 \leq m < n$, $W_m a$ is left special. Let W_{m+1} be the shortest bispecial factor containing the letter a. Thus, $m + 1 \leq n$ since W_n contains a. Since W_m does not contain a, we can write $W_{m+1} = W_m X a Y W_m$. Here any one of X, Y, and W_m may be the empty word. Since W_{m+1} is the shortest bispecial factor containing a, it follows that every occurrence of a in ω is preceded by $W_m X$. Since $W_n a$ is a factor, and W_{m+1} is a suffix of W_n, it follows that $W_m X$ is both a prefix and a suffix of W_{m+1}. But this implies that $W_m X$ is bispecial, and since $|W_m X| < |W_{m+1}|$, we deduce that $W_m X = W_m$, in other words, X is empty. Hence $W_m a$ is left special as required. □ □

We observe that Proposition 3.7.9 holds also for (general) θ-Episturmian words, since for any θ-Episturmian word there exists a standard θ-Episturmian word having the same set of factors.

In general Proposition 3.7.9 does not extend to words $\omega \in SW_\theta(N)$ for $N > 0$. For instance, let $\mathbf{t}$ be the Tribonacci word, i.e., the fixed point of the morphism $0 \mapsto 01$, $1 \mapsto 02$ and $2 \mapsto 0$. Let ω be the image of $\mathbf{t}$ under the morphism $0 \mapsto a$, $1 \mapsto bc$, and $2 \mapsto cab$. Let θ be the involutory antimorphism generated by $\theta(a) = a$, and $\theta(b) = c$. Then it is readily verified that $\omega \in SW_\theta(4)$, but $\omega \notin SW_\theta(3)$ since both abc and cab are left special factors. We have that $W_1 = a$, $W_2 = abca$, and $W_3 = abcacababca$. Thus although $W_2 c$ is left special, and c occurs in W_2, we have that $W_3 \neq (W_2 c)^{\oplus_\theta} = abcacbabca$.

3.8 Special factors and images of Arnoux-Rauzy words

The main result of this section shows that, for generalized Episturmian words in the standard case, even when the step of dropping the "closure under some ϑ" requirement is made, the large class of words thus obtained retains a strong link with Arnoux-Rauzy words. More precisely, we will prove the following.

Theorem 3.8.1. *Let $s \in A^\omega$ satisfy the following two conditions for all $k \geq N$, where $N \geq 0$:*

1. *any left special factor of s having length k is a prefix of s,*

2. *s has at most one right special factor of length k.*

Then there exists $B \subseteq \operatorname{alph} s$ and a standard Arnoux-Rauzy word $t \in B^\omega$ such that s is a morphic image (under an injective morphism) of t.

The following simple lemma is the first basic ingredient for our main result.

Lemma 3.8.2. *Let s be an infinite word such that any sufficiently long left special factor of s is a prefix of it. Then s is recurrent.*

Proof. By contradiction, suppose that λw is a prefix of s ending with the rightmost occurrence of w in s. Then all prefixes of s from length $|\lambda w|$ on do not reoccur in s, and so have *no left extensions* in s. By a counting argument, this implies that s has also at least one factor with *more than one* left extension (i.e., a left special factor) for each length $n \geq |\lambda w|$. For sufficiently large n, such a left special factor should be a prefix of s by hypothesis. We have reached a contradiction. □

We need one of the most well-known and useful restatements of the theorem of Morse and Hedlund (cf. [50, Theorem 7.3]):

Theorem 3.8.3. *An infinite word s is ultimately periodic if and only if $c_s(n) = c_s(n+1)$ for some $n \geq 0$.*

As a consequence of Lemma 3.8.2, we obtain the following specialization.

Proposition 3.8.4. *An infinite word s is (purely) periodic if and only if it has no left special factor of some length n.*

Proof. If $s = p^\omega$ with $p \in A^*$, then s has no left special factors of length $|p|$. Conversely, assume that s has no left special factor of length n. This implies

$$\operatorname{card}(A^n \cap \operatorname{Fact} s) = \operatorname{card}(A^{n+1} \cap \operatorname{Fact} s) ,$$

so that by Theorem 3.8.3, s is ultimately periodic. Clearly s has no left special factor of any length $k \geq n$, thus it trivially satisfies the hypothesis of Lemma 3.8.2. Therefore s is recurrent, and hence periodic. □

The following proposition was proved in [14, Lemma 7] under different hypotheses. We report an adapted proof for the sake of completeness.

Proposition 3.8.5. *Let s be a recurrent aperiodic infinite word. Then every factor w of s is contained in some bispecial factor of s.*

Proof. Since s is recurrent, we can consider a complete return z to w in s. Writing $z = vw$, it cannot happen that the factor w is always preceded by v in s, otherwise s would be periodic. Thus some suffix of z of length at least $|u|$ must be a left special factor of s. Let $x \in A^*$ be of minimal length such that xw is a left special factor of s. Such a word is trivially unique, and w is always preceded in s by x. In a similar way, there exists a unique $y \in A^*$ of minimal length such that wy is right special in s, and w is always followed by y.

Since xw is left special in s and xw is always followed by y one has that xwy is also left special. Similarly, since wy is right special and always preceded by x, xwy is right special. Hence every factor w of s is contained in some bispecial factor $W = xwy$ of s. □

A recurrent word $s \in A^\omega$ with $A = \operatorname{alph} s$ is an *Arnoux-Rauzy* word if it has exactly one left special factor and one right special factor of each length, of degree $\operatorname{card} A$. Arnoux-Rauzy words are uniformly recurrent (cf. [29]); this was part of the definition in [4]. An Arnoux-Rauzy word s is *standard* if its left special factors are prefixes of s. Thanks to Lemma 3.8.2, one does not have to consider recurrence, when checking if a given word is a standard Arnoux-Rauzy word.

Proof of Theorem 3.8.1

Suppose first that s has no left special factor of some length n. Then s is periodic by Proposition 3.8.4, so that it is trivially a morphic image of x^ω for any $x \in \operatorname{alph} s$.

Now let us assume that s has at least one left special factor of each length — exactly one, from length N on. By Lemma 3.8.2, s is recurrent, so that by Proposition 3.8.5 it has infinitely many bispecial factors, which we denote by $W_0 = \varepsilon, W_1, \ldots, W_n, \ldots$, where $|W_i| \leq |W_{i+1}|$ for all $i \geq 0$. Let j be the least index such that $|W_j| \geq N$. By conditions 1 and 2, W_i is a border of W_{i+1}

for all $i \geq j$. The sequence whose n-th term is the (right) degree of W_n for all $n \geq j$ is then non-increasing. Hence there exists $k \geq j$ such that W_n has the same degree of W_k for all $n \geq k$, that is, the above considered sequence is constant from its k-th term on. We set

$$B = \{x \in A \mid W_k x \in \mathrm{Fact}\, s\} \subseteq \mathrm{alph}\, s\ ,$$

so that card$B$ is, by definition, the degree of W_k.

We now consider the return words to $w = W_k$ in s. Let $u_1 w = w v_1$ and $u_2 w = w v_2$ be any two distinct complete returns to w in s, and let us show that $v_1^f \neq v_2^f$. Indeed, let p be the longest common prefix of v_1 and v_2. If $p = v_1$, then $|v_2| > |v_1|$ as $v_1 \neq v_2$; since $wv_1 = u_1 w$, there is an internal occurrence of w in wv_2, contradicting the definition of complete return. The same argument applies if $p = v_2$. Thus p is a proper prefix of both v_1 and v_2, so that wp is a right special factor of s. Since s has only one right special factor per length, and w is a right special factor of s, it follows that w is a suffix of wp. This implies $p = \varepsilon$, since otherwise there would be an internal occurrence of w in wv_1 and wv_2. Hence $v_1^f \neq v_2^f$ as desired. Since w is also left special in s, using a symmetric argument one can prove that $u_1^\ell \neq u_2^\ell$.

From this it follows that for each $x \in B$, there exists a unique complete return $u_x w = w v_x$ to w in s, such that $v_x^f = x$. We define a morphism $\phi : B^* \to A^*$ by $\phi(x) = u_x$. Note that ϕ is injective, as $\phi(B)$ is a suffix code having the same cardinality of B.

By definition, we have $s = \phi(t)$, where $t \in B^\omega$ is a derivated word of s with respect to its prefix w. We remark that, as a consequence of the definition of return words, one has

$$z \in \mathrm{Fact}\, t \Leftrightarrow \phi(z)w \in \mathrm{Fact}\, s \quad \text{ and } \quad z \in \mathrm{Pref}\, t \Leftrightarrow \phi(z)w \in \mathrm{Pref}\, s\ . \qquad (3.22)$$

We will prove that t is a standard Arnoux-Rauzy word by showing that t has exactly one right special factor (of degree card$B$) of each length, and that all left special factors of t are prefixes of it.

Let z_1 and z_2 be any two right special factors of t having the same length. Thus there exist distinct letters $x_1, y_1, x_2, y_2 \in B$ such that $x_i \neq y_i$ and $z_i x_i, z_i y_i \in \mathrm{Fact}\, t$ for $i = 1, 2$. By (3.22), this implies $\phi(z_i x_i)w, \phi(z_i y_i)w \in \mathrm{Fact}\, s$. Since for $\alpha \in \{x_i, y_i\}$ and $i = 1, 2$ we have

$$\phi(z_i \alpha)w = \phi(z_i) u_\alpha w = \phi(z_i) w v_\alpha \in \mathrm{Fact}\, s$$

with $v^f_{x_i} \neq v^f_{y_i}$, it follows that $\phi(z_1)w$ and $\phi(z_2)w$ are right special factors of s. By condition 2, either $\phi(z_1)w \in \mathrm{Suff}(\phi(z_2)w)$, or vice versa. The word w has $|z_1| + 1 = |z_2| + 1$ occurrences in both $\phi(z_1)w$ and $\phi(z_2)w$, and it is a prefix of both, by the definition of return word. Hence we derive $\phi(z_1)w = \phi(z_2)w$, so that $z_1 = z_2$ by the injectivity of ϕ.

If z is a right special factor of t, by the above argument $\phi(z)w$ is right special in s. Since $|\phi(z)w| \geq |w|$, we obtain that $\phi(z)wx \in \mathrm{Fact}\, s$ for all $x \in B$. Since the only complete return to w in s starting with x is v_x, it follows $\phi(z)wv_x = \phi(z)u_x w = \phi(zx)w \in \mathrm{Fact}\, s$, so that $zx \in \mathrm{Fact}\, t$ for all $x \in B$, proving that z has degree $\mathrm{card} B$.

Let now z' be a left special factor of t, and let $xz', yz' \in \mathrm{Fact}\, t$ for some distinct letters $x, y \in B$. Then $\phi(xz')w, \phi(yz')w \in \mathrm{Fact}\, s$. As $\phi(x)^\ell = u^\ell_x \neq u^\ell_y = \phi(y)^\ell$, $\phi(z')w$ is a left special factor of s. By condition 1, it follows $\phi(z')w \in \mathrm{Pref}\, s$ and then $z' \in \mathrm{Pref}\, t$ by (3.22).

Chapter 4

Characteristic morphisms

4.1 Basic definitions and properties

Let X be a finite alphabet. A morphism $\phi : X^* \to A^*$ will be called ϑ-*characteristic* if

$$\phi(SEpi(X)) \subseteq SEpi_\vartheta \ ,$$

i.e., ϕ maps any standard Episturmian word over the alphabet X in a standard ϑ-Episturmian word on the alphabet A. Following this terminology, Theorem 2.3.1 can be reformulated by saying that *an injective morphism* $\phi : A^* \to A^*$ *is standard Episturmian if and only if it is* R*-characteristic.*

For instance, every morphism $\phi : X^* \to A^*$ satisfying the conditions of Proposition 3.6.7 is ϑ-characteristic (and injective). A trivial example of a non-injective ϑ-characteristic morphism is the constant morphism $\phi : x \in X \mapsto a \in A$, where a is a fixed ϑ-palindromic letter.

Let $X = \{x, y\}$, $A = \{a, b, c\}$, ϑ defined by $\bar{a} = a$, $\bar{b} = c$, and $\phi : X^* \to A^*$ be the injective morphism such that $\phi(x) = a$, $\phi(y) = bac$. If t is any standard Episturmian word beginning in y^2x, then $s = \phi(t)$ begins with $bacbaca$, so that a is a left special factor of s which is not a prefix of s. Thus s is not ϑ-Episturmian and therefore ϕ is not ϑ-characteristic.

In this section we shall prove some results concerning the structure of ϑ-characteristic morphisms.

Proposition 4.1.1. *Let* $\phi : X^* \to A^*$ *be a* ϑ*-characteristic morphism. For each* x *in* X*,* $\phi(x) \in PAL_\vartheta^2$*.*

Proof. It is clear that $|\phi(x)|$ is a period of each prefix of $\phi(x^\omega)$. Since $\phi(x^\omega)$ is in $SEpi_\vartheta$, it has infinitely many ϑ-palindromic prefixes (see [12]). Then, from Lemma 3.1.5 the statement follows. □

Let $\phi : X^* \to A^*$ be a morphism such that $\phi(X) \subseteq \mathcal{P}_\vartheta^*$. For any $x \in X$, let $\phi(x) = \pi_1^{(x)} \cdots \pi_{r_x}^{(x)}$ be the unique factorization of $\phi(x)$ by the elements of $\mathcal{P}_\vartheta$. We set

$$\Pi(\phi) = \{\pi \in \mathcal{P}_\vartheta \mid \exists x \in X,\, \exists i :\ 1 \leq i \leq r_x \text{ and } \pi = \pi_i^{(x)}\}\ . \tag{4.1}$$

If ϕ is a ϑ-characteristic morphism, then by Propositions 4.1.1 and 3.6.2, we have $\phi(X) \subseteq PAL_\vartheta^2 \subseteq \mathcal{P}_\vartheta^*$, so that $\Pi(\phi)$ is well defined.

Proposition 4.1.2. *Let $\phi : X^* \to A^*$ be a ϑ-characteristic morphism. Then $\Pi(\phi)$ is an overlap-free and normal code.*

Proof. Let $t \in SEpi(X)$ be such that $\operatorname{alph} t = X$, and consider $s = \phi(t) \in SEpi_\vartheta$. Then the set $\Pi(\phi)$ equals Π_s, as defined in (3.18). The result follows from Theorems 3.4.9 and 3.4.10. □

Proposition 4.1.3. *Let $\phi : X^* \to A^*$ be a ϑ-characteristic morphism. If there exist two letters $x, y \in X$ such that $\phi(x)^f \neq \phi(y)^f$, then $\phi(X) \subseteq PAL_\vartheta$.*

Proof. Set $w = \phi((x^2y)^\omega)$. Clearly $\phi(x)$ is a right special factor of w, since it appears followed both by $\phi(x)$ and $\phi(y)$. As w is in $SEpi_\vartheta$, being the image of the standard Episturmian word $(x^2y)^\omega$, we have that $\overline{\phi(x)}$ is a left special factor, and thus a prefix, of w. But also $\phi(x)$ is a prefix of w, then it must be $\phi(x) = \overline{\phi(x)}$. The same argument can be applied to $\phi(y)$, setting $w' = \phi((y^2x)^\omega)$.

Now let $z \in X$. Then $\phi(z)^f$ cannot be equal to both $\phi(x)^f$ and $\phi(y)^f$. Therefore, as shown above, $\phi(z) \in PAL_\vartheta$. From this the assertion follows. □

Proposition 4.1.4. *Let $\phi : X^* \to A^*$ be a ϑ-characteristic morphism. If for $x, y \in X$, $\operatorname{Suff}\phi(x) \cap \operatorname{Suff}\phi(y) \neq \{\varepsilon\}$, then $\phi(xy) = \phi(yx)$, that is, both $\phi(x)$ and $\phi(y)$ are powers of a word of A^*.*

Proof. If $\phi(xy) \neq \phi(yx)$, since $\operatorname{Suff}\phi(x) \cap \operatorname{Suff}\phi(y) \neq \{\varepsilon\}$, there exists a common proper suffix h of $\phi(xy)$ and $\phi(yx)$, with $h \neq \varepsilon$. Let h be the longest of such suffixes. Then there exist $v, u \in A^+$ such that

$$\phi(xy) = vh \quad \text{and} \quad \phi(yx) = uh \; , \tag{4.2}$$

with $v^\ell \neq u^\ell$. Let s be a standard Episturmian word whose directive word can be written as $\Delta = xy^2x\lambda$, with $\lambda \in X^\omega$, so that $s = xyxyxxyxyxt$, with $t \in X^\omega$. Thus

$$\phi(s) = \phi(xy)\phi(xy)\alpha = \phi(x)\phi(yx)\phi(yx)\phi(xy)\beta$$

for some $\alpha, \beta \in A^\omega$. By (4.2), it follows

$$\phi(s) = v\underline{hv}h\alpha = \phi(x)uhu\underline{hv}h\beta \; .$$

The underlined occurrences of hv are preceded by different letters, namely v^ℓ and u^ℓ. Since $\phi(s) \in SEpi_\vartheta$, this implies $hv \in \operatorname{Pref} \phi(s)$ and then

$$hv = vh \; . \tag{4.3}$$

In a perfectly symmetric way, by considering an Episturmian word s' whose directive word Δ' has yx^2y as a prefix, we obtain that $uh = hu$. Hence u and h are powers of a common primitive word w; by (4.3), the same can be said about v and h. Since the primitive root of a nonempty word is unique, it follows that u and v are both powers of w. As $|u| = |v|$ by definition, we obtain $u = v$ and then $\phi(xy) = \phi(yx)$, which is a contradiction. □

Corollary 4.1.5. *If $\phi : X^* \to A^*$ is an injective ϑ-characteristic morphism, then $\phi(X)$ is a suffix code.*

Proof. It is clear that if ϕ is injective, then for all $x, y \in X, x \neq y$, one has $\phi(xy) \neq \phi(yx)$; from Proposition 4.1.4 it follows $\operatorname{Suff} \phi(x) \cap \operatorname{Suff} \phi(y) = \{\varepsilon\}$. Thus, for all $x, y \in X$, if $x \neq y$, then $\phi(x) \notin \operatorname{Suff} \phi(y)$, and the statement follows. □

Proposition 4.1.6. *Let $\phi : X^* \to A^*$ be a ϑ-characteristic morphism. Then for each $x, y \in X$, either*

$$\operatorname{alph} \phi(x) \cap \operatorname{alph} \phi(y) = \emptyset$$

or

$$\phi(x)^f = \phi(y)^f .$$

Proof. Let $\mathrm{alph}\,\phi(x) \cap \mathrm{alph}\,\phi(y) \neq \emptyset$ and $\phi(x)^f \neq \phi(y)^f$. We set p as the longest prefix of $\phi(x)$ such that $\mathrm{alph}\,p \cap \mathrm{alph}\,\phi(y) = \emptyset$ and $c \in A$ such that $pc \in \mathrm{Pref}\,\phi(x)$. Let then p' be the longest prefix of $\phi(y)$ in which c does not appear, i.e., such that $c \notin \mathrm{alph}\,p'$. Since we have assumed that $\phi(x)^f \neq \phi(y)^f$, it cannot be $p = p' = \varepsilon$. Let us suppose that both $p \neq \varepsilon$ and $p' \neq \varepsilon$. In this case we have that c is left special in $(\phi(xy))^\omega$, since it appears preceded both by p and p' and, from the definition of p, $\mathrm{alph}\,p \cap \mathrm{alph}\,p' = \emptyset$. We reach a contradiction, since c should be a prefix of $\phi(xy)^\omega$ which is in $SEpi_\vartheta$, and thus a prefix of $\phi(x)$.

We then have that either $p \neq \varepsilon$ and $p' = \varepsilon$ or $p = \varepsilon$ and $p' \neq \varepsilon$. In the first case we set $z = x$ and $z' = y$, otherwise we set $z' = x$ and $z = y$. Thus we can write

$$\phi(z) = \lambda c \gamma, \quad \phi(z') = c\gamma' , \tag{4.4}$$

with $\lambda \in A^+$, $c \notin \mathrm{alph}\,\lambda$, and $\gamma, \gamma' \in A^*$. For each nonnegative integer n, $(z^n z')^\omega$ and $(z'^n z)^\omega$ are standard Episturmian words, so that $(\phi(z^n z'))^\omega$ and $(\phi(z'^n z))^\omega$ are in $SEpi_\vartheta$. Moreover, since

$$(\phi(zz'))^\omega = \phi(z')^{-1}(\phi(z'z))^\omega \quad \text{and} \quad (\phi(z'z))^\omega = \phi(z)^{-1}(\phi(zz'))^\omega ,$$

it is clear that $\phi(zz')^\omega$ and $\phi(z'z)^\omega$ have the same set of factors, so that each left special factor of $(\phi(zz'))^\omega$ is a left special factor of $(\phi(z'z))^\omega$ and *vice versa*.

Let w be a nonempty left special factor of $(\phi(z'z))^\omega$; then w is also a prefix. As noted above, w has to be a left special factor (and thus a prefix) of $(\phi(zz'))^\omega$. Thus w is a common prefix of $(\phi(z'z))^\omega$ and $(\phi(zz'))^\omega$, which is a contradiction since the first word begins with c whereas the second begins with λ, which does not contain c. Therefore $\phi(z'z)^\omega$ has no left special factor different from ε; since each right special factor of a word in $SEpi_\vartheta$ is the ϑ-image of a left special factor, it is clear that $(\phi(z'z))^\omega$ has no special factor different from ε.

Hence each factor of $(\phi(z'z))^\omega$ can be extended in a unique way both to the left and to the right, so that by (4.4) we can write

$$(\phi(z'z))^\omega = c\gamma'\lambda c \cdots$$

and, as stated above, each occurrence of c must be followed by $\gamma'\lambda c$, which yields that

$$(\phi(z'z))^\omega = (c\gamma'\lambda)^\omega = (\phi(z')\lambda)^\omega ,$$

so that this infinite word has the two periods $|\phi(z'z)|$ and $|\phi(z')\lambda|$. From the theorem of Fine and Wilf, one derives $\phi(z'z)(\phi(z')\lambda) = (\phi(z')\lambda)\phi(z'z)$, so that

$$\phi(zz')\lambda = \lambda\phi(z'z) \ . \tag{4.5}$$

The preceding equation tells us that λ is a suffix of $\lambda\phi(z'z)$ and so, as $|\phi(z)| > |\lambda|$, it must be a suffix of $\phi(z)$; since λ does not contain any c, it has to be a suffix of γ, so that we can write

$$\phi(z) = \lambda cg\lambda \tag{4.6}$$

for some word g. Substituting in (4.5), it follows

$$\phi(zz') = \lambda\phi(z')\lambda cg \ .$$

From the preceding equation, we have

$$(\phi(z'^2z))^\omega = \phi(z')\phi(z')\lambda\phi(z')\lambda cg\cdots \tag{4.7}$$

From (4.6), $\phi(z)^\ell = \lambda^\ell$. Proposition 4.1.4 ensures that $\lambda^\ell = \phi(z)^\ell$ must be different from $\phi(z')^\ell$, otherwise we would obtain $\phi(zz') = \phi(z'z)$ which would imply c is a prefix of $\phi(z)$, which is a contradiction. Thus, from (4.7), we have that $\phi(z')\lambda$ is a left special factor of $\phi(z'^2z)^\omega$ and this implies that $\phi(z')\lambda$ is a prefix of $\phi(z')^2\phi(z)$, from which we obtain that λ is a prefix of $\phi(z'z) = c\gamma'\phi(z)$, that is a contradiction, since λ does not contain any occurrence of c. Thus the initial assumption that $\mathrm{alph}\,\phi(x) \cap \mathrm{alph}\,\phi(y) \neq \emptyset$ and $\phi(x)^f \neq \phi(y)^f$, leads in any case to a contradiction. $\square$

Proposition 4.1.7. *Let $\phi : X^* \to A^*$ be a ϑ-characteristic morphism. If $x, y \in X$ and $\phi(x), \phi(y) \in PAL_\vartheta$, then either* $\mathrm{alph}\,\phi(x) \cap \mathrm{alph}\,\phi(y) = \emptyset$ *or $\phi(xy) = \phi(yx)$. In particular, if ϕ is injective and $\phi(X) \subseteq PAL_\vartheta$, then for all $x, y \in X$ with $x \neq y$ we have* $\mathrm{alph}\,\phi(x) \cap \mathrm{alph}\,\phi(y) = \emptyset$.

Proof. If $\mathrm{alph}\,\phi(x) \cap \mathrm{alph}\,\phi(y) \neq \emptyset$, from Proposition 4.1.6 we obtain, as $\phi(x), \phi(y) \in PAL_\vartheta$, that $\overline{\phi(x)^\ell} = \phi(x)^f = \phi(y)^f = \overline{\phi(y)^\ell}$. Then $\phi(x)^\ell = \phi(y)^\ell$ and, from Proposition 4.1.4, we have that $\phi(xy) = \phi(yx)$.

If ϕ is injective, then for all $x, y \in X$ with $x \neq y$ we have $\phi(xy) \neq \phi(yx)$ so that the assertion follows. $\square$

Corollary 4.1.8. *Let $\phi : X^* \to A^*$ be an injective ϑ-characteristic morphism such that $\phi(X) \subseteq PAL_\vartheta$ and card$X \geq 2$. Then $\phi(X) \subseteq \mathcal{P}_\vartheta$.*

Proof. Let $x, y \in X$ with $x \neq y$. Since ϕ is injective, we have from Proposition 4.1.7 that $\operatorname{alph}\phi(x) \cap \operatorname{alph}\phi(y) = \emptyset$. Let u be a proper border of $\phi(x)$. Then there exist two nonempty words v and w such that

$$\phi(x) = uv = wu.$$

Since $\operatorname{alph}\phi(x) \cap \operatorname{alph}\phi(y) = \emptyset$, we have $\phi(y)^\ell \neq w^\ell$; thus

$$\phi(yx)^\omega = \phi(y)uv\phi(y)wu\phi(y)\cdots$$

shows that u is a left special factor in $\phi(yx)^\omega$, but this would imply that u is a prefix of $\phi(yx)$. As $\operatorname{alph} u \cap \operatorname{alph}\phi(y) = \emptyset$, it follows $u = \varepsilon$, i.e., $\phi(x) \in \mathcal{P}_\vartheta$. The same argument applies to $\phi(y)$. □

The following lemma will be useful in the next section.

Lemma 4.1.9. *Let $\phi : X^* \to A^*$ be a ϑ-characteristic morphism. Then for each $x \in X$ and for any $a \in A$,*

$$|\phi(x)|_a > 1 \Longrightarrow |\phi(x)|_{\phi(x)^f} > 1.$$

Proof. Let b be the first letter of $\phi(x)$ such that $|\phi(x)|_b > 1$. Then we can write

$$\phi(x) = vbwbw'$$

with $w, w' \in A^*$, $b \notin \operatorname{alph} w$, and $|\phi(x)|_c = 1$ for each c in $\operatorname{alph} v$. If $v \neq \varepsilon$, then we have that $v^\ell \neq (bw)^\ell$, but that means that b is left special in $\phi(x^\omega)$, which is a contradiction, since each left special factor of $\phi(x^\omega)$ is a prefix and b is not in $\operatorname{alph} v$. Then it must be $v = \varepsilon$ and $b = \phi(x)^f$. □

4.2 First results

The first result of this section is a characterization of injective ϑ-characteristic morphisms such that the image of any letter is an unbordered ϑ-palindrome. A wider characterization will be given in Section 4.3.

Theorem 4.2.1. *Let $\phi : X^* \to A^*$ be an injective morphism such that for any $x \in X$, $\phi(x) \in \mathcal{P}_\vartheta$. Then ϕ is ϑ-characteristic if and only if the following two conditions hold:*

1. $\operatorname{alph}\phi(x) \cap \operatorname{alph}\phi(y) = \emptyset$, *for any x, y in X such that $x \neq y$.*

2. *for any $x \in X$ and $a \in A$, $|\phi(x)|_a \leq 1$.*

Proof. Let ϕ be ϑ-characteristic. Since ϕ is injective, from Proposition 4.1.7 we have that if $x \neq y$, then $\operatorname{alph}\phi(x) \cap \operatorname{alph}\phi(y) = \emptyset$. Thus condition 1 holds. Let us now prove that condition 2 is satisfied. This is certainly true if $|\phi(x)| \leq 2$, as $\phi(x) \in \mathcal{P}_\vartheta$. Let us then suppose $|\phi(x)| > 2$. We can write

$$\phi(x) = ax_1 \cdots x_n b \ ,$$

with $x_i \in A$, $i = 1, \ldots, n$, $\bar{a} = b$, and $a \neq b$.

Let us prove that for any $i = 1, \ldots, n$, $x_i \notin \{a, b\}$. By contradiction, suppose that b has an internal occurrence in $\phi(x)$, and consider its first occurrence. Since $\phi(x)$ is a ϑ-palindrome, we can write

$$\phi(x) = ax_1 \cdots x_i b\lambda = \bar{\lambda} a \bar{x}_i \cdots \bar{x}_1 b \ ,$$

with $\lambda \in A^*$, $1 \leq i < n$, and $x_j \neq b$ for $j = 1, \ldots, i$.

We now consider the standard ϑ-Episturmian word $s = \phi(x^\omega)$, whose first letter is a. We have that no letter $\bar{x}_j$, $j = 1, \ldots, i$, is left special in s, as otherwise $\bar{x}_j = a$ that implies $x_j = b$, which is absurd. Also b cannot be left special since otherwise $b = a$. Thus it follows that $x_i = \bar{x}_1$, $x_{i-1} = \bar{x}_2$, $\ldots$, $x_1 = \bar{x}_i$. Hence, $ax_1 \cdots x_i b$ is a proper border of $\phi(x)$, which is a contradiction. From this, since $\phi(x)$ is a ϑ-palindrome, one derives that there is no internal occurrence of a in $\phi(x)$ as well.

Finally, any letter of $\phi(x)$ cannot occur more than once. This is a consequence of Lemma 4.1.9, since otherwise the first letter of $\phi(x)$, namely a, would reoccur in $\phi(x)$. Thus condition 2 holds.

Conversely, let us now suppose that conditions 1 and 2 hold; Proposition 3.6.7 ensures then that ϕ is ϑ-characteristic. $\square$

A different proof of Theorem 4.2.1 will be given at the end of this section, as a consequence of a full characterization of injective ϑ-characteristic morphisms, given in Theorem 4.3.1.

Remark. In the "if" part of Theorem 4.2.1 the requirement $\phi(X) \subseteq \mathcal{P}_\vartheta$ can be replaced by $\phi(X) \subseteq PAL_\vartheta$, as condition 2 implies that $\phi(x)$ is unbordered for any $x \in X$, so that $\phi(X) \subseteq \mathcal{P}_\vartheta$. In the "only if" part, in view of Corollary 4.1.8, one can replace $\phi(X) \subseteq \mathcal{P}_\vartheta$ by $\phi(X) \subseteq PAL_\vartheta$ under the hypothesis that $\mathrm{card}X \geq 2$.

Example 4.2.2. Let X, A, ϑ, and g be defined as in Example 3.4.6. Then the morphism g is ϑ-characteristic.

As an immediate consequence of Theorem 4.2.1, we obtain:

Corollary 4.2.3. *Let $\zeta : X^* \to B^*$ be an R-characteristic morphism, $g : B^* \to A^*$ be an injective morphism satisfying $g(B) \subseteq \mathcal{P}_\vartheta$ and the two conditions in the statement of Theorem 4.2.1. Then $\phi = g \circ \zeta$ is ϑ-characteristic.*

Example 4.2.4. Let X, A, ϑ, and g be defined as in Example 3.4.6, and let ζ be the endomorphism of X^* such that $\zeta(x) = xy$ and $\zeta(y) = xyx$. Since $\zeta = \mu_{xy} \circ \sigma$, where $\sigma(x) = y$ and $\sigma(y) = x$, ζ is a standard Episturmian morphism. Hence the morphism $\phi : X^* \to A^*$ given by

$$\phi(x) = acbde, \quad \phi(y) = acbdeacb$$

is ϑ-characteristic, as $\phi = g \circ \zeta$.

Theorem 4.2.5. *Let $\phi : X^* \to A^*$ be a ϑ-characteristic morphism. Then there exist $B \subseteq A$, a morphism $\zeta : X^* \to B^*$, and a morphism $g : B^* \to A^*$ such that:*

1. *ζ is R-characteristic,*
2. *$g(B) = \Pi(\phi)$, with $g(b) \in bA^*$ for all $b \in B$,*
3. *$\phi = g \circ \zeta$.*

Proof (see Fig. 4.1). Set $\Pi = \Pi(\phi)$, as defined in (4.1), and let $B = f(\Pi) \subseteq A$, where f is the morphism considered in (3.2). Let $\phi_| : X^* \to \Pi^*$ and $f_| : \Pi^* \to B^*$ be the restrictions of ϕ and f, respectively. Setting $\zeta = f_| \circ \phi_| : X^* \to B^*$, by Theorem 3.4.11 one derives $\zeta(SEpi(X)) \subseteq SEpi(B)$, i.e., ζ is R-characteristic.

Let $t \in SEpi(X)$ be such that $\mathrm{alph}\, t = X$, and consider $s = \phi(t) \in SEpi_\vartheta$. Since Π equals Π_s, as defined in (3.18), by Theorem 3.4.11 the morphism f is

$$\begin{tikzcd}
X^* \arrow[rr, "\phi"] \arrow[dr, "\phi_{|}"] \arrow[ddr, "\zeta"'] & & A^* \\
& \Pi^* \arrow[ur, "\iota"'] \arrow[d, shift left, "f_{|}"] & \\
& B^* \arrow[u, dashed, shift left] \arrow[uur, "g"'] &
\end{tikzcd}$$

Figure 4.1: A commutative diagram describing Theorem 4.2.5

injective over Π, so that $f_|$ is bijective. Set $g = \iota \circ f_|^{-1}$, where $\iota : \Pi^* \to A^*$ is the inclusion map. Then $g(B) = \Pi$, and $g(b) \in bA^*$ for all $b \in B$. Furthermore, we have

$$\phi = \iota \circ \phi_| = \iota \circ (f_|^{-1} \circ f_|) \circ \phi_| = (\iota \circ f_|^{-1}) \circ (f_| \circ \phi_|) = g \circ \zeta$$

as desired. □

Example 4.2.6. Let $X = \{x, y\}$, $A = \{a, b, c\}$, and ϑ be the antimorphism of A^* such that $\bar{a} = a$ and $\bar{b} = c$. The morphism $\phi : X^* \to A^*$ defined by $\phi(x) = a$ and $\phi(y) = abac$ is ϑ-characteristic (this will be clear after Theorem 4.3.1, see Example 4.3.2), and it can be decomposed as $\phi = g \circ \zeta$, where $\zeta : X^* \to B^*$ (with $B = \{a, b\}$) is the morphism such that $\zeta(x) = a$ and $\zeta(y) = ab$, while $g : B^* \to A^*$ is defined by $g(a) = a$ and $g(b) = bac$. We remark that $\zeta(SEpi(X)) \subseteq SEpi(B)$, but $g(SEpi(B)) \not\subseteq SEpi_\vartheta$ as it can be verified using Theorem 4.2.1. Observe that this example shows that not all ϑ-characteristic morphisms can be constructed as in Corollary 4.2.3.

Proposition 4.2.7. *Let $\zeta : X^* \to A^*$ be an injective morphism. Then ζ is R-characteristic if and only if it can be decomposed as $\zeta = \mu_w \circ \eta$, where $w \in A^*$ and $\eta : X^* \to A^*$ is an injective literal morphism.*

Proof. Let $\zeta = \mu_w \circ \eta$, with $w \in A^*$ and η an injective literal morphism. Then η is trivially R-characteristic and μ_w is R-characteristic too, by Theorem 2.3.1. Therefore also their composition ζ is R-characteristic.

Conversely, let us first suppose that $\zeta(X) \subseteq a_1 A^*$ for some $a_1 \in A$. Then for any $t \in SEpi(X)$, $\zeta(t)$ is a standard Episturmian word beginning with a_1, so that by Proposition 2.2.9 the letter a_1 is separating for $\zeta(t)$. In particular a_1 is separating for each $\zeta(x)$ ($x \in X$); by Proposition 2.2.15 there exists a morphism $\alpha_1 : X^* \to A^*$ such that $\zeta = \mu_{a_1} \circ \alpha_1$. Since $t \in SEpi(X)$, $\mu_{a_1}(\alpha_1(t))$ is a standard Episturmian word over A, so that by Proposition 2.3.2 the word $\alpha_1(t)$ is also a standard Episturmian word over A. Thus α_1 is injective and R-characteristic, and we can iterate the above argument to find new letters $a_i \in A$ and R-characteristic morphisms α_i such that $\zeta = \mu_{a_1} \circ \cdots \circ \mu_{a_i} \circ \alpha_i$, as long as all images of letters under α_i have the same first letter.

If $\mathrm{card} X > 1$, since ζ is injective, we eventually obtain the following decomposition:

$$\zeta = \mu_{a_1} \circ \mu_{a_2} \circ \cdots \circ \mu_{a_n} \circ \eta = \mu_w \circ \eta \ , \tag{4.8}$$

where $a_1, \ldots, a_n \in A$, $w = a_1 \cdots a_n$, and $\eta = \alpha_n$ is such that $\eta(x)^f \neq \eta(y)^f$ for some $x, y \in X$. If the original requirement $\zeta(X) \subseteq a_1 A^*$ is not met by any a_1, that is, if $\zeta(x)^f \neq \zeta(y)^f$ for some $x, y \in X$, we can still fit in (4.8) choosing $n = 0$ and $w = \varepsilon$.

Let then $x, y \in X$ be such that $\eta(x)^f \neq \eta(y)^f$. Since η is R-characteristic, by Proposition 4.1.3 we obtain $\eta(X) \subseteq PAL$. Moreover, since η is injective, by Corollary 4.1.8 we have $\eta(X) \subseteq \mathcal{P}_R = A$, so that η is an injective literal morphism.

In the case $X = \{x\}$, the lengths of the words $\alpha_i(x)$ for $i \geq 1$ are decreasing. Hence eventually we find an $n \geq 1$ such that $\alpha_n(x) \in A$ and the assertion is proved, for

$$\zeta = \mu_{a_1} \circ \cdots \circ \mu_{a_n} \circ \alpha_n = \mu_w \circ \alpha_n \ ,$$

with $w = a_1 \cdots a_n \in A^*$ and $\alpha_n : X^* \to A^*$ an injective literal morphism. $\square$

Example 4.2.8. Let $X = \{x, y\}$, $A = \{a, b, c\}$, and $\zeta : X^* \to A^*$ be defined by:

$$\zeta(x) = abacabaabacab = \mu_a(bcbabcb) \quad \text{and} \quad \zeta(y) = abacaba = \mu_a(bcba) \ ,$$

so that $\alpha_1(x) = bcbabcb$ and $\alpha_1(y) = bcba$. Then $\zeta(x)$ can be rewritten also as

$$\zeta(x) = \mu_a(\alpha_1(x)) = (\mu_a \circ \mu_b)(cacb) = (\mu_a \circ \mu_b \circ \mu_c)(ab) = \mu_{abca}(b) \ .$$

In a similar way, one obtains $\zeta(y) = \mu_{abca}(a)$. Hence, setting $\eta(x) = b$ and $\eta(y) = a$, the morphism $\zeta = \mu_{abca} \circ \eta$ is R-characteristic, in view of the preceding proposition.

From Theorem 4.2.5 and Proposition 4.2.7 one derives the following:

Corollary 4.2.9. *Every injective ϑ-characteristic morphism $\phi : X^* \to A^*$ can be decomposed as*

$$\phi = g \circ \mu_w \circ \eta \ , \tag{4.9}$$

where $\eta : X^* \to B^*$ is an injective literal morphism, $\mu_w : B^* \to B^*$ is a pure standard Episturmian morphism (with $w \in B^*$), and $g : B^* \to A^*$ is an injective morphism such that $g(B) = \Pi(\phi)$.

Remarks.

1. From the preceding result, we have in particular that if $\phi : X^* \to A^*$ is an injective ϑ-characteristic morphism, then $\mathrm{card} X \leq \mathrm{card} A$.

2. Theorem 4.2.5 and Proposition 4.2.7 show that a decomposition (4.9) can always be chosen so that $B = \mathrm{alph}\, w \cup \eta(X) \subseteq A$ and $g(b) \in bA^* \cap \mathcal{P}_\vartheta$ for each $b \in B$.

3. Corollary 4.2.9 shows that the code $\phi(X)$, which is a suffix code by Corollary 4.1.5, is in fact the *composition* (by means of g) [7] of the code $\mu_w(\eta(X)) \subseteq B^*$ and the biprefix, overlap-free, and normal code $g(B) \subseteq A^*$.

4. From the proof of Proposition 4.2.7, one easily obtains that if $\mathrm{card} X > 1$, the decomposition (4.9) is unique.

Proposition 4.2.10. *Let $\phi : X^* \to A^*$ be an injective ϑ-characteristic morphism, decomposed as in* (4.9)*. The word $u = g(\psi(w))$ is a ϑ-palindrome such that for each $x \in X$,*

$$\phi(x)u = (u\, g(\eta(x)))^{\oplus} \ , \tag{4.10}$$

and $\phi(x)$ is either a prefix of u or equal to $ug(\eta(x))$.

Proof. Since $\psi(w)$ is a palindrome and the injective morphism g is such that $g(B) \subseteq \mathcal{P}_\vartheta$, we have $u \in PAL_\vartheta$ in view of Proposition 3.1.9. Let $x \in X$ and set $b = \eta(x)$. We have

$$\phi(x)u = g(\mu_w(\eta(x))\psi(w)) = g(\mu_w(b)\psi(w)).$$

By Propositions 2.2.13 and 3.1.9 we obtain

$$g(\mu_w(b)\psi(w)) = g(\psi(wb)) = g((\psi(w)b)^{(+)}) = (g(\psi(w)b))^{\oplus} = (ug(b))^{\oplus},$$

and (4.10) follows. Thus, since $g(b)$ is a ϑ-palindromic suffix of $ug(b)$, we derive $|\phi(x)| \leq |ug(b)|$. By Proposition 4.1.1, $\phi(x) \in \mathcal{P}_\vartheta^*$. Therefore it can be either equal to $ug(b)$ or a prefix of u. Indeed, if $\phi(x) = ur$ with r a nonempty proper prefix of $g(b) \in \mathcal{P}_\vartheta$, then $r \in \mathcal{P}_\vartheta^*$, as $\mathcal{P}_\vartheta^*$ is left unitary. This gives rise to a contradiction because $\mathcal{P}_\vartheta$ is a biprefix code. □

Corollary 4.2.11. *Under the same hypotheses and with the same notation as in Proposition 4.2.10, if $x_1, x_2 \in X$ are such that $|\phi(x_1)| \leq |\phi(x_2)|$, then either $\phi(x_1) \in \operatorname{Pref} \phi(x_2)$, or $\phi(x_1)$ and $\phi(x_2)$ do not overlap, i.e.,*

$$\operatorname{Suff} \phi(x_1) \cap \operatorname{Pref} \phi(x_2) = \operatorname{Suff} \phi(x_2) \cap \operatorname{Pref} \phi(x_1) = \{\varepsilon\} .$$

Proof. For $i = 1, 2$, let us set $b_i = \eta(x_i)$. By Proposition 4.2.10, $\phi(x_i)$ is either a prefix of u or equal to $ug(b_i)$.

If $\phi(x_1)$ is a prefix of u, then it is a prefix of $\phi(x_2)$ too, as $|\phi(x_1)| \leq |\phi(x_2)|$. Let us then suppose that

$$\phi(x_i) = ug(b_i) \quad \text{for } i = 1, 2 . \tag{4.11}$$

Now let v be an element of $\operatorname{Suff} \phi(x_1) \cap \operatorname{Pref} \phi(x_2)$. Since $\phi(x_2) \in \mathcal{P}_\vartheta^*$, we can write $v = v'\lambda$, where v' is the longest word of $\mathcal{P}_\vartheta^* \cap \operatorname{Pref} v$. Then λ is a proper prefix of a word π occurring in the unique factorization of $\phi(x_2)$ over $\mathcal{P}_\vartheta$. If λ were nonempty, π would overlap with some word π' of the factorization of $\phi(x_1)$ over $\mathcal{P}_\vartheta$. This is absurd, since for any $t \in SEpi(X)$ such that $x_1, x_2 \in \operatorname{alph} t$, both π and π' would be in $\Pi_{\phi(t)}$, which is overlap-free by Theorem 3.4.10. Hence $\lambda = \varepsilon$ and $v \in \mathcal{P}_\vartheta^*$. Therefore by (4.11) we have $v = g(\xi)$, where ξ is an element of $\operatorname{Suff}(\psi(w)b_1) \cap \operatorname{Pref}(\psi(w)b_2)$.

By Proposition 4.2.10, (4.11) is equivalent to $(u\,g(b_i))^{\oplus} = ug(b_i)u$, $i = 1, 2$. Since for $i = 1, 2$ the word $g(b_i)$ is an unbordered ϑ-palindrome, any ϑ-palindromic suffix of $ug(b_i)$ longer than $g(b_i)$ can be written as $g(b_i)\xi_i g(b_i)$, with ξ_i a ϑ-palindromic suffix of u. Hence (4.11) holds for $i = 1, 2$ if and only if u has no ϑ-palindromic suffixes preceded respectively by $g(b_1)$ or $g(b_2)$. By Proposition 3.1.9, this implies that for $i = 1, 2$, $\psi(w)$ has no palindromic suffix preceded by b_i, so that $b_i \notin \operatorname{alph} w = \operatorname{alph} \psi(w)$. Therefore, since $b_1 \neq b_2$, the only word in $\operatorname{Suff}(\psi(w)b_1) \cap \operatorname{Pref}(\psi(w)b_2)$ is ε. Hence $v = g(\varepsilon) = \varepsilon$.

The same argument can be used to prove that $\operatorname{Suff} \phi(x_2) \cap \operatorname{Pref} \phi(x_1) = \{\varepsilon\}$. □

Example 4.2.12. Let $X = \{x, y\}$, $A = \{a, b, c, d, e\}$, $B = \{a, d\}$, and ϑ be defined by $\bar{a} = b$, $\bar{c} = c$, and $\bar{d} = e$. As we have seen in Example 4.2.4, the morphism $\phi : X^* \to A^*$ defined by $\phi(x) = acbde$ and $\phi(y) = acbdeacb$ is ϑ-characteristic. We can decompose ϕ as $\phi = g \circ \mu_{ad} \circ \eta$, where $g : B^* \to A^*$ is defined by $g(a) = acb \in \mathcal{P}_\vartheta$, $g(d) = de \in \mathcal{P}_\vartheta$, and η is such that $\eta(x) = d$ and $\eta(y) = a$. We have $u = g(\psi(ad)) = g(ada) = acbdeacb$, and

$$\phi(x)u = acbdeacbdeacb = (acbdeacbde)^\oplus = (u\, g(\eta(x)))^\oplus \ .$$

Similarly, $\phi(y)u = (u\, g(\eta(y)))^\oplus$. In this case, $\phi(x)$ is a prefix of $\phi(y)$.

4.3 A characterization of characteristic morphisms

The following basic theorem gives a characterization of all injective ϑ-characteristic morphisms.

Theorem 4.3.1. *Let $\phi : X^* \to A^*$ be an injective morphism. Then ϕ is ϑ-characteristic if and only if it is decomposable as*

$$\phi = g \circ \mu_w \circ \eta$$

as in (4.9), *with $B = \operatorname{alph} w \cup \eta(X)$ and $g(B) = \Pi \subseteq \mathcal{P}_\vartheta$ satisfying the following conditions:*

1. *Π is an overlap-free and normal code,*
2. *$LS\left(\{g(\psi(w))\} \cup \Pi\right) \subseteq \operatorname{Pref} g(\psi(w))$,*
3. *if $b, c \in A \setminus \operatorname{Suff} \Pi$ and $v \in \Pi^*$ are such that $bv\bar{c} \in \operatorname{Fact} \Pi$, then $v = g(\psi(w'x))$, with $w' \in \operatorname{Pref} w$ and $x \in \{\varepsilon\} \cup (B \setminus \eta(X))$.*

The proof of this theorem, which is rather cumbersome, will be given at the end of this section, using some results on biprefix, overlap-free, and normal codes proved in Section 1.3. Before presenting the proof, we give some examples and a remark related to Theorem 4.3.1; furthermore, we derive from this theorem a different proof of Theorem 4.2.1.

Example 4.3.2. Let $A = \{a, b, c\}$, $X = \{x, y\}$, $B = \{a, b\}$, and let ϑ and $\phi : X^* \to A^*$ be defined as in Example 4.2.6, namely $\bar{a} = a$, $\bar{b} = c$, and $\phi = g \circ \mu_a \circ \eta$, where $\eta(x) = a$, $\eta(y) = b$, and $g : B^* \to A^*$ is defined by $g(a) = a$ and $g(b) = bac$. Then $\Pi = g(B) = \{a, bac\}$ is an overlap-free code and satisfies:

- $(\operatorname{Suff}\Pi \setminus \Pi) \cap LS\,\Pi = \{\varepsilon\}$, so that Π is normal,
- $LS(\{g(\psi(a))\} \cup \Pi) = LS(\{a\} \cup \Pi) = \{\varepsilon\} \subseteq \operatorname{Pref} a$.

The only word verifying the hypotheses of condition 3 is $bac = ba\bar{b} = g(b) \in \Pi$, with $a \in \Pi^*$ and $b \notin \operatorname{Suff}\Pi$. Since $a = g(\psi(a))$ and $B \setminus \eta(X) = \emptyset$, also condition 3 of Theorem 4.3.1 is satisfied. Hence ϕ is ϑ-characteristic.

Example 4.3.3. Let $X = \{x, y\}$, $A = \{a, b, c\}$, ϑ be such that $\bar{a} = a$, $\bar{b} = c$, and the morphism $\phi : X^* \to A^*$ be defined by $\phi(x) = a$ and $\phi(y) = abaac$. In this case we have $\phi = g \circ \mu_a \circ \eta$, where $B = \{a, b\}$, $g(a) = a$, $g(b) = baac$, $\eta(x) = a$, and $\eta(y) = b$. Then the morphism ϕ is not ϑ-characteristic. Indeed, if t is any standard Episturmian word starting with yxy, then $\phi(t)$ has the prefix $abaacaabaac$, so that aa is a left special factor of $\phi(t)$ but not a prefix of it.

In fact, condition 3 of Theorem 4.3.1 is not satisfied in this case, since $baac = baa\bar{b} = g(b)$, $b \notin \operatorname{Suff}\Pi$, $aa \in \Pi^*$, $B \setminus \eta(X) = \emptyset$, and

$$aa \notin \{g(\psi(w')) \mid w' \in \operatorname{Pref} a\} = \{\varepsilon, a\}\ .$$

If we choose $X' = \{y\}$ with $\eta'(y) = b$, then

$$g(\mu_a(\eta'(y^\omega))) = (abaac)^\omega \in SEpi_\vartheta\ ,$$

so that $\phi' = g \circ \mu_a \circ \eta'$ is ϑ-characteristic. In this case $B = \operatorname{alph} a \cup \eta'(X')$, $B \setminus \eta'(X') = \{a\}$, and $aa = g(\psi(aa)) = g(aa)$, so that condition 3 is satisfied.

Example 4.3.4. Let $X = \{x, y\}$, $A = \{a, b, c, d, e, h\}$, and ϑ be the antimorphism over A defined by $\bar{a} = a$, $\bar{b} = c$, $\bar{d} = e$, $\bar{h} = h$. Let also $w = adb \in A^*$, $B = \{a, b, d\} = \operatorname{alph} w$, and $\eta : X^* \to B^*$ be defined by $\eta(x) = a$ and $\eta(y) = b$. Finally, set $g(a) = a$, $g(d) = dahae$, and $g(b) = badahaeadahaeac$. Then the morphism $\phi = g \circ \mu_w \circ \eta$ is such that

$$\phi(y) = adahaeabadahaeadahaeac \quad \text{and} \quad \phi(x) = \phi(y)\, adahaea\,,$$

and it is ϑ-characteristic as the code $\Pi = g(B)$ and the word $u = g(\psi(w)) = g(adabada) = \phi(x)$ satisfy all three conditions of Theorem 4.3.1.

Remark. Let us observe that Theorem 4.3.1 gives an effective procedure to decide whether, for a given ϑ, an injective morphism $\varphi : X^* \to A^*$ is ϑ-characteristic. The procedure runs in the following steps:

1. Check whether $\varphi(X) \subseteq \mathcal{P}_\vartheta^*$.

2. If the previous condition is satisfied, then compute $\Pi = \Pi(\varphi)$.

3. Verify that Π is overlap-free and normal.

4. Compute $B = f(\Pi)$ and then the morphism $g : B^* \to A^*$ given by $g(B) = \Pi$.

5. Since $\varphi = g \circ \zeta$, verify that ζ is R-characteristic, i.e., there exists $w \in B^*$ such that $\zeta = \mu_w \circ \eta$, where η is a literal morphism from X^* to B^*. This can be always simply done, following the argument used in the proof of Proposition 4.2.7.

6. Compute $g(\psi(w))$ and verify that conditions 2 and 3 of Theorem 4.3.1 are satisfied. This can also be effectively done.

Before going on, we now give a new proof of Theorem 4.2.1, based on Theorem 4.3.1.

Proof of Theorem 4.2.1. Let $\phi : X^* \to A^*$ be an injective morphism such that $\phi(X) = \Pi \subseteq \mathcal{P}_\vartheta$ and satisfying conditions 1 and 2 of Theorem 4.2.1. In this case we can assume $w = \varepsilon$, so that $B = \eta(X)$, $u = g(\psi(w)) = \varepsilon$, and $\phi = g \circ \eta$. Hence $\Pi = g(B) = \phi(X)$. The code Π is overlap-free by conditions 1 and 2. Since any letter of A occurs at most once in any word of Π, we have $LS(\{\varepsilon\} \cup \Pi) \subseteq \{\varepsilon\} = \operatorname{Pref} u$, whence

$$(\operatorname{Suff} \Pi \setminus \Pi) \cap LS\,\Pi \subseteq \{\varepsilon\} ,$$

i.e., Π is a left normal, and therefore normal, code. Let $b, c \in A \setminus \operatorname{Suff} \Pi$, and $v \in \Pi^*$ be such that $bv\bar{c} \in \operatorname{Fact} \pi$ for some $\pi \in \Pi$. This implies $v = \varepsilon = g(\psi(\varepsilon))$, because the equation $v = \pi_1 \cdots \pi_k$ with $\pi_1, \ldots, \pi_k \in \Pi$ would violate condition 1 of Theorem 4.2.1. Thus all the hypotheses of Theorem 4.3.1 are satisfied for $w = \varepsilon$, so that $\phi = g \circ \mu_\varepsilon \circ \eta$ is ϑ-characteristic.

Conversely, let $\phi : X^* \to A^*$ be an injective ϑ-characteristic morphism such that $\phi(X) = \Pi \subseteq \mathcal{P}_\vartheta$. We can take $w = \varepsilon$, $B = \eta(X) \subseteq A$ and write $\phi = g \circ \eta$, so that $g(B) = \phi(X) = \Pi$. Since $u = \varepsilon$, by Theorem 4.3.1 we have

$$LS(\{\varepsilon\} \cup \Pi) \subseteq \{\varepsilon\} , \tag{4.12}$$

and, as $B \setminus \eta(X) = \emptyset$, for all $b, c \in A \setminus \operatorname{Suff} \Pi$ and $v \in \Pi^*$,

$$bv\bar{c} \in \operatorname{Fact} \Pi \implies v = g(\psi(\varepsilon)) = \varepsilon . \tag{4.13}$$

Moreover, since $\Pi = \Pi(\phi)$, we have that Π is normal and overlap-free by Proposition 4.1.2.

Now let $a \in A$ and suppose $a \in \operatorname{alph} \pi$ for some $\pi \in \Pi$. We will show that *any* two occurrences of a in the words of Π coincide, so that a has exactly one occurrence in Π. Let then $\pi_1, \pi_2 \in \Pi$ be such that

$$\pi_1 = \lambda_1 a \rho_1 \quad \text{and} \quad \pi_2 = \lambda_2 a \rho_2$$

for some $\lambda_1, \lambda_2, \rho_1, \rho_2 \in A^*$, and let us first prove that $\lambda_1 = \lambda_2$.

Let s be the longest common suffix of λ_1 and λ_2, and let $\lambda_i = \lambda_i' s$ for $i = 1, 2$. If both λ_1' and λ_2' were nonempty, their last letters would differ by the definition of s, and therefore sa would be in $LS\,\Pi$, contradicting (4.12).

Next, we may assume $\lambda_1' = \varepsilon$ and $\lambda_2' \neq \varepsilon$, without loss of generality. Then $sa \in \operatorname{Pref} \pi_1$, so that by Proposition 1.3.1 we obtain $\lambda_2' \pi_1 \in \operatorname{Pref} \pi_2$; in particular, we have $\pi_1 \neq \pi_2$. Let then r be the longest word of $\Pi^* \cap \operatorname{Suff} \lambda_2'$, and set $\lambda_2' = \xi r$. Since $\lambda_2' \neq \varepsilon$ and Π is a biprefix code, we have $\xi \neq \varepsilon$. Furthermore, ξ^ℓ is not a suffix of any word of Π, for if π' were such a word, by Proposition 1.3.1 we would derive that $\pi' \in \operatorname{Suff} \xi$, contradicting the definition of r.

Let us now write $\pi_2 = \xi r \pi_1 \delta$. The word δ is nonempty since Π is a biprefix code. Let r' be the longest word in $\Pi^* \cap \operatorname{Pref}\delta$ and set $\delta = r'\zeta$. Since Π is a biprefix code, $\zeta \neq \varepsilon$. By Proposition 1.3.1, we derive that $\zeta^f \notin \operatorname{Pref}\Pi$. By (4.13), we obtain that $r\pi_1 r' = \varepsilon$, which is absurd.

Thus $\lambda_1' = \lambda_2' = \varepsilon$, whence $\lambda_1 = \lambda_2$ as desired. From $\lambda_1 a = \lambda_2 a$ it follows $\pi_1^f = \pi_2^f$, so that by Proposition 1.3.1 we have $\pi_1 = \pi_2$ and hence $\rho_1 = \rho_2$. Therefore, the two (generic) occurrences of a we have considered are the same.

We have thus proved that every letter of A occurs at most once among all the words of $\Pi = \phi(X)$, so that conditions 1 and 2 of Theorem 4.2.1 are satisfied. $\square$

In order to prove theorem 4.3.1, we need the following lemma.

Lemma 4.3.5. *Let $t \in SEpi(B)$ with $\operatorname{alph} t = B$, and let $s = g(t)$ be a standard ϑ-Episturmian word over A, with $g : B^* \to A^*$ an injective morphism such that $g(B) \subseteq \mathcal{P}_\vartheta$. Suppose that $b, c \in A \setminus \operatorname{Suff}\Pi_s$ and $v \in \Pi_s^*$ are such that $bv\bar{c} \in \operatorname{Fact}\Pi_s$. Then there exists $\delta \in B^*$ such that $v = g(\psi(\delta))$.*

Proof. Let $\pi \in \Pi_s$ be such that $bv\bar{c} \in \operatorname{Fact}\pi$. By definition, we have $\Pi_s = g(B)$, so that, since $v \in \Pi_s^*$, we can write $v = g(\xi)$ for some $\xi \in B^*$. We have to prove that $\xi = \psi(\delta)$ for some $\delta \in B^*$. This is trivial for $\xi = \varepsilon$. Let then $\psi(\delta')$ be the longest prefix in $\psi(B^*)$ of ξ, and assume by contradiction that $\xi \neq \psi(\delta')$, so that $\psi(\delta')a \in \operatorname{Pref}\xi$ for some $a \in B$. We shall prove that $\psi(\delta' a) = (\psi(\delta')a)^{(+)} \in \operatorname{Pref}\xi$, contradicting the maximality of $\psi(\delta')$.

Since $g(\psi(\delta'))$ is a prefix of v, we have $bg(\psi(\delta')) \in \operatorname{Fact}\pi \subseteq \operatorname{Fact}s$. Moreover $g(\psi(\delta')a) \in \operatorname{Pref}v \subseteq \operatorname{Fact}\pi$. By Proposition 3.1.9 and since π is a ϑ-palindrome, we have

$$g(a\psi(\delta')) = \overline{g(\psi(\delta')a)} \in \operatorname{Fact}\pi \ .$$

Thus $g(\psi(\delta'))$, being preceded in s both by $b \notin \operatorname{Suff}\Pi_s$ and by $(g(a))^\ell \in \operatorname{Suff}\Pi_s$, is a left special factor of s, and hence a prefix of it.

Suppose first that $a \notin \operatorname{alph}\delta'$, so that $\psi(\delta' a) = \psi(\delta')a\psi(\delta')$. Let λ be the longest prefix of $\psi(\delta')$ such that $\psi(\delta')a\lambda$ is a prefix of ξ. Then $g(\psi(\delta')a\lambda)$ is followed in $v\bar{c}$ by some letter x, i.e.,

$$g(\psi(\delta')a\lambda)x \in \operatorname{Pref}(v\bar{c}) \ . \tag{4.14}$$

We claim that

$$g(\lambda)x \notin \operatorname{Pref} g(\psi(\delta')) . \tag{4.15}$$

Indeed, assume the contrary. Then x is a prefix of $g(\lambda)^{-1}g(\psi(\delta'))$, which is in Π^* since Π is a biprefix code. Hence $x \in \operatorname{Pref} g(d)$ for some $d \in B$ such that $g(\lambda d) \in \operatorname{Pref} g(\psi(\delta'))$, and then $\lambda d \in \operatorname{Pref} \psi(\delta')$ by Lemma 1.3.2. As $\bar{c} \notin \operatorname{Pref} \Pi$, we obtain $x \neq \bar{c}$, so that by (4.14) it follows $g(\psi(\delta')a\lambda)x \in \operatorname{Pref} v$. Therefore $g(\psi(\delta')a\lambda d) \in \operatorname{Pref} v$ by Proposition 1.3.1, so that $\psi(\delta')a\lambda d \in \operatorname{Pref} \xi$ by Lemma 1.3.2. This is a contradiction because of our choice of λ.

Let us prove that $\lambda = \psi(\delta')$. Indeed, since $\tilde{\lambda} \in \operatorname{Suff} \psi(\delta')$, by (4.14) the word $g(\tilde{\lambda}a\lambda)x$ is a factor of π, and so is its image under ϑ, that is $\bar{x}g(\tilde{\lambda}a\lambda)$. By contradiction, suppose $|\lambda| < |\psi(\delta')|$. By (4.15), $\bar{x}g(\tilde{\lambda}) \notin \operatorname{Suff} g(\psi(\delta'))$, so that the suffix $g(\tilde{\lambda}a\lambda)$ of $g(\psi(\delta')a\lambda)$ is preceded by a letter which is not $\bar{x}$. Thus $g(\tilde{\lambda}a\lambda)$ is a left special factor of $\pi \in \operatorname{Fact} s$, and hence a prefix of s. As we have previously seen, $g(\psi(\delta'))$ is a prefix of s too, so that, as $|\lambda| < |\psi(\delta')|$, it follows by Lemma 1.3.2 that $\tilde{\lambda}a$ is a prefix of $\psi(\delta')$, contradicting the hypothesis that $a \notin \operatorname{alph} \delta'$. Thus $\lambda = \psi(\delta')$, so that $\psi(\delta'a) \in \operatorname{Pref} \xi$, as we claimed.

Now let us assume $a \in \operatorname{alph} \delta'$ instead, and write $\delta' = \gamma a \gamma'$ with $a \notin \operatorname{alph} \gamma'$, so that $\psi(\delta') = \psi(\gamma)a\rho = \tilde{\rho}a\psi(\gamma)$ and $\psi(\gamma)$ is the longest palindromic prefix (resp. suffix) of $\psi(\delta')$ followed (resp. preceded) by a. Thus

$$\psi(\delta'a) = \tilde{\rho}a\psi(\gamma)a\rho = \psi(\delta')a\rho .$$

Let $\lambda \in \operatorname{Pref} \rho$ and $x \in A$ be such that (4.14) holds and $g(\lambda)x \notin \operatorname{Pref} g(\rho)$. With the same argument as above, one can show that if $|\lambda| < |\rho|$, then $g(\tilde{\lambda}a\psi(\gamma)a\lambda)$ is a left special factor, and then a prefix, of s. Since $g(\psi(\delta'))$ is a prefix of s too, and $|\tilde{\lambda}a\psi(\gamma)a| \leq |\rho a\psi(\gamma)| = |\psi(\delta')|$, by Lemma 1.3.2 we obtain $\tilde{\lambda}a\psi(\gamma)a \in \operatorname{Pref} \psi(\delta')$. Since $\tilde{\lambda}$ is a suffix of $\tilde{\rho}$, $\tilde{\lambda}a\psi(\gamma)$ is a suffix, and then a border, of $\psi(\delta')$. This is absurd since $\psi(\gamma)$ is the longest border of $\psi(\delta')$ followed by a. Thus $\lambda = \rho$, showing that $\psi(\delta'a)$ is a prefix of ξ also in this case. The proof is complete. $\square$

We can now proceed with the proof of Theorem 4.3.1.

Necessity. The decomposition (4.9) with $B = \operatorname{alph} w \cup \eta(X)$ follows from Corollary 4.2.9 and subsequent Remark.

Since $\Pi = g(B) \subseteq \mathcal{P}_\vartheta$ and ϕ is ϑ-characteristic, one has by Theorem 4.2.5 that $\Pi = \Pi(\phi)$ as defined by (4.1), so that it is overlap-free and normal by Proposition 4.1.2.

Let us set $u = g(\psi(w))$, and prove that condition 2 holds. We first suppose that $\operatorname{card} X \geq 2$, and that $a, a' \in \eta(X)$ are distinct letters. Let Δ be an infinite word such that $\operatorname{alph}\Delta = \eta(X)$. Setting $t_a = \psi(wa\Delta)$ and $t_{a'} = \psi(wa'\Delta)$, by (2.12) we have

$$t_a = \mu_w(\psi(a\Delta)) \quad \text{and} \quad t_{a'} = \mu_w(\psi(a'\Delta)) ,$$

so that, setting $s_y = g(t_y)$ for $y \in \{a, a'\}$, we obtain

$$s_y = g(\mu_w(\psi(y\Delta))) \in SEpi_\vartheta$$

as $\psi(y\Delta) \in \eta(SEpi(X)) \subseteq SEpi(B)$ and $\phi = g \circ \mu_w \circ \eta$ is ϑ-characteristic. By Corollary 2.2.14 and (2.12), one obtains that the longest common prefix of t_a and $t_{a'}$ is $\psi(w)$. As $\operatorname{alph}\Delta = \eta(X)$ and $B = \operatorname{alph} w \cup \eta(X)$, we have $\operatorname{alph} t_a = \operatorname{alph} t_{a'} = B$, so that $\Pi_{s_a} = \Pi_{s_{a'}} = \Pi$. Since g is injective, by Theorem 3.4.11 we have $g(a)^f \neq g(a')^f$, so that the longest common prefix of s_a and $s_{a'}$ is $u = g(\psi(w))$. Any word of $LS(\{u\} \cup \Pi)$, being a left special factor of both s_a and $s_{a'}$, has to be a common prefix of s_a and $s_{a'}$, and hence a prefix of u.

Now let us suppose $X = \{z\}$ and denote $\eta(z)$ by a. In this case we have

$$\phi(SEpi(X)) = \{g(\mu_w(a^\omega))\} = \{(g(\mu_w(a)))^\omega\} .$$

Let us set $s = (g(\mu_w(a)))^\omega \in SEpi_\vartheta$. By Corollary 2.2.14, $u = g(\psi(w))$ is a prefix of s. Let $\lambda \in LS(\{u\} \cup \Pi)$. Since $\Pi = \Pi_s$, the word λ is a left special factor of the ϑ-Episturmian word s, so that we have $\lambda \in \operatorname{Pref} s$.

If $a \in \operatorname{alph} w$, then $B = \{a\} \cup \operatorname{alph} w = \operatorname{alph} w = \operatorname{alph} \psi(w)$, so that $\Pi \subseteq \operatorname{Fact} u$. This implies $|\lambda| \leq |u|$ and then $\lambda \in \operatorname{Pref} u$ as desired.

If $a \notin \operatorname{alph} w$, then by Proposition 4.2.10 we obtain $\phi(z) = g(\mu_w(a)) = u\, g(a)$, because $\phi(z) \notin \operatorname{Pref} u$ otherwise by Lemma 1.3.2 we would obtain $\mu_w(a) \in \operatorname{Pref} \psi(w)$, that implies $a \in \operatorname{alph} w$. Hence $s = (u\, g(a))^\omega$. Since $\Pi \subseteq \{g(a)\} \cup \operatorname{Fact} u$, we have $|\lambda| \leq |u\, g(a)|$, so that $\lambda \in \operatorname{Pref}(u\, g(a))$. Again, if λ is a proper prefix of u we are done, so let us suppose that $\lambda = u\lambda'$ for some $\lambda' \in \operatorname{Pref} g(a)$, and that λ is a left special factor of $g(a)$. Then the prefix λ' of

$g(a)$ is repeated in $g(a)$. The longest repeated prefix p of $g(a)$ is either a right special factor or a border of $g(a)$. Both possibilities imply $p = \varepsilon$, since $g(a)$ is unbordered and Π is a biprefix and normal code. As $\lambda' \in \operatorname{Pref} p$, it follows $\lambda' = \varepsilon$. This proves condition 2.

Finally, let us prove condition 3. Let $b, c \in A \setminus \operatorname{Suff} \Pi$, $v \in \Pi^*$, and $\pi \in \Pi$ be such that $bv\bar{c} \in \operatorname{Fact} \pi$. Let $t' \in SEpi(X)$ with $\operatorname{alph} t' = X$, and set $t = \mu_w(\eta(t'))$, $s_1 = g(t)$. Since ϕ is ϑ-characteristic, $s_1 = \phi(t')$ is standard ϑ-Episturmian. By Lemma 4.3.5, we have $v = g(\psi(\delta))$ for some $\delta \in B^*$. If $\delta = \varepsilon$ we are done, as condition 3 is trivially satisfied for $w' = x = \varepsilon$; let us then write $\delta = \delta' a$ for some $a \in B$. The words $bg(\psi(\delta'))$ and $g(a\psi(\delta'))$ are both factors of the ϑ-palindrome π; indeed, $\psi(\delta' a)$ begins with $\psi(\delta')a$ and terminates with $a\psi(\delta')$. Hence $g(\psi(\delta'))$ is left special in π as $b \notin \operatorname{Suff} \Pi$ is different from $(g(a))^\ell \in \operatorname{Suff} \Pi$. Therefore $g(\psi(\delta'))$ is a prefix of $g(\psi(w))$, as we have already proved condition 2. Since g is injective and Π is a biprefix code, by Lemma 1.3.2 it follows $\psi(\delta') \in \operatorname{Pref} \psi(w)$, so that $\delta' \in \operatorname{Pref} w$ by Proposition 2.2.10. Hence, we can write $\delta = w'x$ with $w' \in \operatorname{Pref} w$ and x either equal to a (if $\delta' a \notin \operatorname{Pref} w$) or to ε. It remains to show that if $w'x \notin \operatorname{Pref} w$, then $x \notin \eta(X)$.

Let us first assume that $\eta(X) = \{x\}$. In this case we have $s_1 = g(\mu_w(\eta(t'))) = g(\psi(wx^\omega))$ by (2.12). Since $bv = bg(\psi(w'x)) \in \operatorname{Fact} \pi$, $g(x)$ is a proper factor of π. Then, as $B = \{x\} \cup \operatorname{alph} w$ and $g(x) \neq \pi$, we must have $\pi \in g(\operatorname{alph} w)$, so that $bv \in \operatorname{Fact} g(\psi(w))$ as $\operatorname{alph} w = \operatorname{alph} \psi(w)$. By Proposition 2.2.11, $\psi(w'x)$ is a factor of $\psi(wx)$. We can then write $\psi(wx) = \zeta\psi(w'x)\zeta'$ for some $\zeta, \zeta' \in B^*$. If ζ were empty, by Proposition 2.2.10 we obtain $w'x \in \operatorname{Pref}(wx)$. Since $w'x \notin \operatorname{Pref} w$ we would derive $w = w'$, which is a contradiction since we proved that $bv = bg(\psi(w'x)) \in \operatorname{Fact} g(\psi(w))$. Therefore $\zeta \neq \varepsilon$, and v is left special in s, being preceded both by $(g(\zeta))^\ell$ and by $b \notin \operatorname{Suff} \Pi$. This implies that v is a prefix of s and then of $g(\psi(w))$ as $|v| \leq |g(\psi(w))|$. By Lemma 1.3.2, it follows $\psi(w'x) \in \operatorname{Pref} \psi(w)$ and then $w'x \in \operatorname{Pref} w$ by Proposition 2.2.10, which is a contradiction.

Suppose now that there exists $y \in \eta(X) \setminus \{x\}$, and let $\Delta \in \eta(X)^\omega$ with $\operatorname{alph} \Delta = \eta(X)$. The word $s_2 = g(\psi(wyx\Delta))$ is equal to $g(\mu_w(\psi(yx\Delta)))$ by (2.12), and is then standard ϑ-Episturmian since $\phi = g \circ \mu_w \circ \eta$ is ϑ-characteristic. By applying Proposition 2.2.11 to w' and $wy \in w'A^*$, we obtain $\psi(w'x) \in \operatorname{Fact}\psi(wyx)$. We can write $\psi(wyx) = \zeta\psi(w'x)\zeta'$ for some $\zeta, \zeta' \in B^*$. As $w'x \notin \operatorname{Pref} w$ and $x \neq y$, we have by Proposition 2.2.10 that $\psi(w'x) \notin \operatorname{Pref}\psi(wy)$, so that $\zeta \neq \varepsilon$. Hence $v = g(\psi(w'x))$ is left special in s_2, being preceded both by $(g(\zeta))^\ell$ and by $b \notin \operatorname{Suff}\Pi$. This implies that v is a prefix of s_2 and then of $g(\psi(wy))$; by Lemma 1.3.2, this is absurd since $\psi(w'x) \notin \operatorname{Pref}\psi(wy)$. □

Sufficiency. Let $t' \in SEpi(\eta(X))$ and $t = \mu_w(t') \in SEpi(B)$. Since $g(B) = \Pi \subseteq \mathcal{P}_\vartheta$, by Proposition 3.1.9 it follows that $g(t)$ has infinitely many ϑ-palindromic prefixes, so that it is closed under ϑ.

Thus, in order to prove that $g(t) \in SEpi_\vartheta$, it is sufficient to show that any nonempty left special factor λ of $g(t)$ is in $\operatorname{Pref} g(t)$. Since λ is left special, there exist $a, a' \in A$, $a \neq a'$, $v, v' \in A^*$, and $r, r' \in A^\omega$, such that

$$g(t) = va\lambda r = v'a'\lambda r' \;. \tag{4.16}$$

The word $g(t)$ can be uniquely factorized by the elements of Π. Therefore, $va\lambda$ and $v'a'\lambda$ are in $\operatorname{Pref}\Pi^*$. We consider three different cases.

Case 1: $va \notin \Pi^*$, $v'a' \notin \Pi^*$.

Since Π is a biprefix (as it is a subset of $\mathcal{P}_\vartheta$), overlap-free, and normal code, by Proposition 1.3.4 we have $a\lambda, a'\lambda \in \operatorname{Fact}\Pi$. Therefore, by condition 2 of Theorem 4.3.1, it follows $\lambda \in LS\,\Pi \subseteq \operatorname{Pref} g(\psi(w))$, so that it is a prefix of $g(t)$ since by Corollary 2.2.14, $\psi(w)$ is a prefix of $t = \mu_w(t')$.

Case 2: $va \in \Pi^*$, $v'a' \in \Pi^*$.

From (4.16), we have $\lambda \in \operatorname{Pref}\Pi^*$. By Proposition 1.3.5, there exists a unique word $\lambda' \in \Pi^*$ such that $\lambda' = \pi_1 \cdots \pi_k = \lambda\zeta$ and $\pi_1 \cdots \pi_{k-1}\delta = \lambda$, with $k \geq 1$, $\pi_i \in \Pi$ for $i = 1, \ldots, k$, $\delta \in A^+$, and $\zeta \in A^*$.

Since g is injective, there exist and are unique the words $\tau, \gamma, \gamma' \in B^*$ such that $g(\tau) = \lambda', g(\gamma) = va, g(\gamma') = v'a'$. Moreover, we have $g(\gamma\tau) = va\lambda' = va\lambda\zeta \in \operatorname{Pref} g(t)$ and $g(\gamma'\tau) = v'a'\lambda' = v'a'\lambda\zeta \in \operatorname{Pref} g(t)$. By Lemma 1.3.2, we derive $\gamma\tau, \gamma'\tau \in \operatorname{Pref} t$. Setting $\alpha = \gamma^\ell, \alpha' = \gamma'^\ell$, we obtain $\alpha\tau, \alpha'\tau \in \operatorname{Fact} t$, and $\alpha \neq \alpha'$ as $a \neq a'$. Hence τ is a left special factor of t; since $t \in SEpi(B)$, we have $\tau \in \operatorname{Pref} t$, so that $g(\tau) = \lambda' \in \operatorname{Pref} g(t)$. As λ is a prefix of λ', it follows $\lambda \in \operatorname{Pref} g(t)$.

Case 3: $va \notin \Pi^*$, $v'a' \in \Pi^*$ (resp. $va \in \Pi^*$, $v'a' \notin \Pi^*$).

We shall consider only the case when $va \notin \Pi^*$ and $v'a' \in \Pi^*$, as the symmetric case can be similarly dealt with.

Since $v'a' \in \Pi^*$, by (4.16) we have $\lambda \in \operatorname{Pref}\Pi^*$. By Proposition 1.3.5, there exists a unique word $\lambda' \in \Pi^*$ such that $\lambda' = \pi_1 \cdots \pi_k = \lambda\zeta$ and $\pi_1 \cdots \pi_{k-1}\delta = \lambda$, with $k \geq 1$, $\pi_i \in \Pi$ for $i = 1, \ldots, k$, $\delta \in A^+$, and $\zeta \in A^*$. By the uniqueness of λ', $v'a'\lambda'$ is a prefix of $g(t)$.

By (4.16) we have $va\pi_1 \cdots \pi_{k-1}\delta \in \operatorname{Pref} g(t)$. By Proposition 1.3.4, $a\lambda \in \operatorname{Fact}\Pi$, so that there exist $\xi, \xi' \in A^*$, $\pi \in \Pi$, such that

$$\xi a\lambda\xi' = \xi a\pi_1 \cdots \pi_{k-1}\delta\xi' = \pi \in \Pi \,.$$

Since δ is a nonempty prefix of π_k, it follows from Proposition 1.3.1 that $\pi = \xi a\pi_1 \cdots \pi_k\xi'' = \xi a\lambda'\xi''$, with $\xi'' \in A^*$. By Proposition 1.3.6, we can write

$$\pi = \xi a\lambda'\xi'' = hp\lambda'qh'$$

with $h, h' \in A^+$, $p, q \in \Pi^*$, $b = h^\ell \notin \operatorname{Suff}\Pi$, and $\bar{c} = (h')^f \notin \operatorname{Pref}\Pi$.

By condition 3, we have $p\lambda'q = g(\psi(w'x))$ for some $w' \in \operatorname{Pref} w$ and $x \in \{\varepsilon\} \cup (B \setminus \eta(X))$. Since $p, \lambda', q \in \Pi^*$ and g is injective, we derive $\lambda' = g(\tau)$ for some $\tau \in \operatorname{Fact}\psi(w'x)$. We will show that λ' is a prefix of $g(t)$, which proves the assertion as $\lambda \in \operatorname{Pref}\lambda'$.

Suppose first that $p = \varepsilon$, so that $a = b$ and $\tau \in \operatorname{Pref}\psi(w'x)$. If $\tau \in \operatorname{Pref}\psi(w')$, then $\lambda' \in g(\operatorname{Pref}\psi(w')) \subseteq \operatorname{Pref} g(\psi(w')) \subseteq \operatorname{Pref} g(\psi(w))$, and we are done as $g(\psi(w)) \in \operatorname{Pref} g(t)$. Let us then assume $x \neq \varepsilon$, so that $x \in B \setminus \eta(X)$, and $\psi(w')x \in \operatorname{Pref}\tau$. Moreover, we can assume $w'x \notin \operatorname{Pref} w$, for otherwise we would derive $\lambda' \in \operatorname{Pref} g(\psi(w))$ again. Let $\Delta \in \eta(X)^\omega$ be the directive word of t', so that by (2.12) we have $t = \psi(w\Delta)$. Since $w' \in \operatorname{Pref} w$, we can write $w\Delta = w'\Delta'$ for some $\Delta' \in B^\omega$, so that $t = \psi(w'\Delta')$.

We have already observed that $v'a'\lambda' \in \operatorname{Pref} g(t)$; as $v'a' \in \Pi^*$, by Lemma 1.3.2 one derives that τ is a factor of t. Since $\psi(w')x \in \operatorname{Pref}\tau$, it follows $\psi(w')x \in \operatorname{Fact}\psi(w'\Delta')$; by Proposition 2.2.12, we obtain $x \in \operatorname{alph}\Delta'$. This implies, since $x \notin \eta(X)$, that $w \neq w'$, and we can write $w = w'\sigma x\sigma'$ for some $\sigma, \sigma' \in B^*$. By Proposition 2.2.11, $\psi(w'x)$ is a factor of $\psi(w'\sigma x)$ and hence of $\psi(w)$, so that, since $\tau \in \operatorname{Pref}\psi(w'x)$, we have $\tau \in \operatorname{Fact}\psi(w)$. Hence we have either

$\tau \in \operatorname{Pref}\psi(w)$, so that $\lambda' \in \operatorname{Pref} g(\psi(w))$ and we are done, or there exists a letter y such that $y\tau \in \operatorname{Fact}\psi(w)$, so that $d\lambda' \in \operatorname{Fact} g(\psi(w))$ with $d = (g(y))^\ell \in \operatorname{Suff}\Pi$. In the latter case, since $a = b \notin \operatorname{Suff}\Pi$ and $a\lambda' \in \operatorname{Fact}\Pi$, we have by condition 2 that $\lambda' \in \operatorname{Pref} g(\psi(w))$. Since $g(\psi(w))$ is a prefix of $g(t)$, in the case $p = \varepsilon$ the assertion is proved.

If $p \neq \varepsilon$, we have $a \in \operatorname{Suff}\Pi$. Let then $\alpha, \alpha' \in B$ be such that $(g(\alpha))^\ell = a$ and $(g(\alpha'))^\ell = a'$; as $a \neq a'$, we have $\alpha \neq \alpha'$. Since $p\lambda'$ is a prefix of $g(\psi(w'x))$, $p \in \Pi^*$, and $p^\ell = (g(\alpha))^\ell = a$, by Lemma 1.3.2 one derives that $\alpha\tau$ is a factor of $\psi(w'x)$. Moreover, as $v'a'\lambda' \in \operatorname{Pref} g(t)$ and $v'a' \in \Pi^*$, we derive that $\alpha'\tau$ is a factor of t.

Let then δ' be any prefix of the directive word Δ of t', such that $\alpha'\tau \in \operatorname{Fact}\psi(w\delta')$. By Proposition 2.2.11, $\psi(w\delta'x)$ contains $\psi(w'x)$, and hence $\alpha\tau$, as a factor. Thus τ is a left special factor of $\psi(w\delta'x)$ and then of the standard Episturmian word $\psi(w\delta'x^\omega)$; as $|\tau| < |\psi(w\delta')|$, it follows $\tau \in \operatorname{Pref}\psi(w\delta')$ and then $\tau \in \operatorname{Pref} t$, so that $\lambda' \in \operatorname{Pref} g(t)$. The proof is now complete. $\square$

4.4 Further results on characteristic morphisms

Theorem 3.4.11 shows that every standard ϑ-Episturmian word is a morphic image, under a suitable injective morphism, of some standard Episturmian word. The following theorem improves upon this, showing that the morphism can always be taken to be ϑ-characteristic.

Theorem 4.4.1. *Let s be a standard ϑ-Episturmian word over A. Then there exists $X \subseteq A$, $t' \in SEpi(X)$ and an injective ϑ-characteristic morphism $\phi : X^* \to A^*$ such that $s = \phi(t')$.*

Proof. Set $\Pi = \Pi_s$. By Theorem 3.4.11, the restriction to Π of the map $f : w \in \mathcal{P}_\vartheta \mapsto w^f \in A$ is injective. Hence, setting $B = f(\Pi) \subseteq A$, we can define an injective morphism g sending any letter $x \in B$ to the only word of Π beginning with x. We have $s = g(t)$, where $t = f(s) \in SEpi(B)$ by Theorem 3.4.11.

Let now $w \in B^*$ be the longest word such that $\psi(w) \in \operatorname{Pref} t$ and $g(\psi(w)) \in \operatorname{Fact}\Pi$. Such a word certainly exists, as $\varepsilon = \psi(\varepsilon) \in \operatorname{Pref} t$ and $\varepsilon = g(\psi(\varepsilon)) \in \operatorname{Fact}\Pi$. Since $\psi(w) \in \operatorname{Pref} t$, we can write t as $\psi(w\Delta)$ for some $\Delta \in B^\omega$; let us set

$$X = \operatorname{alph}\Delta \subseteq B \quad \text{and} \quad t' = \psi(\Delta) \in SEpi(X)\,.$$

By (2.12) we obtain $s = \phi(t')$, where $\phi = g \circ \mu_w \circ \eta$ and η is the inclusion map of X in B, i.e., $\eta(X) = X$.

Let us now show that ϕ is ϑ-characteristic. We have $B = X \cup \operatorname{alph} w$, and $g(B) = \Pi_s \subseteq \mathcal{P}_\vartheta$ is a biprefix code. By Theorems 3.4.9 and 3.4.10, Π is also normal and overlap-free, so that condition 1 of Theorem 4.3.1 is satisfied.

Let us first prove that ϕ meets condition 3 of that theorem. Indeed, if $v \in \Pi^*$ and $b, c \in A \setminus \operatorname{Suff}\Pi$ are such that $bv\bar{c} \in \operatorname{Fact}\pi$ with $\pi \in \Pi$, then by Lemma 4.3.5 we have $v = g(\psi(\delta))$ for some $\delta \in B^*$. If $\delta = \varepsilon$ we are done; let us then write $\delta = \delta' a$ for some $a \in B$. The words $bg(\psi(\delta'))$ and $g(a\psi(\delta'))$ are both factors of the ϑ-palindrome π, so that $g(\psi(\delta'))$ is left special in π as $b \notin \operatorname{Suff}\Pi$ is different from $(g(a))^\ell$. Therefore $g(\psi(\delta')) \in \operatorname{Pref} g(t)$, so that by Lemma 1.3.2 we have $\psi(\delta') \in \operatorname{Pref} t$. Since $g(\psi(\delta')) \in \operatorname{Fact}\Pi$, from the maximality condition on w it follows $|\delta'| \leq |w|$. Moreover, as $\psi(w) \in \operatorname{Pref} t$, by Proposition 2.2.10 it follows $\delta' \in \operatorname{Pref} w$. Hence, we can write $\delta = w'x$ with $w' \in \operatorname{Pref} w$ and x either equal to a (if $\delta' a \notin \operatorname{Pref} w$) or to ε.

In order to prove condition 3, it remains to show that if $w'x \notin \operatorname{Pref} w$, then $x \notin X$. By contradiction, assume $x \in X = \operatorname{alph}\Delta$ and write $\Delta = \xi x \Delta'$ for some $\xi \in (X \setminus \{x\})^*$ and $\Delta' \in X^\omega$. From (2.12), it follows $t = \psi(w\xi x\Delta')$. By applying Proposition 2.2.11 to w' and $w\xi \in w'B^*$, we obtain $\psi(w'x) \in \operatorname{Fact}\psi(w\xi x)$; let us write $\psi(w\xi x) = \zeta\psi(w'x)\zeta'$ for some $\zeta, \zeta' \in B^*$. We claim that $\zeta \neq \varepsilon$, i.e., $\psi(w'x) \notin \operatorname{Pref}\psi(w\xi x)$. Indeed, assume the contrary. Then $w'x \in \operatorname{Pref}(w\xi x)$ by Proposition 2.2.10, so that $w' = w$ and $\xi = \varepsilon$ since $w'x \notin \operatorname{Pref} w$ and $x \notin \operatorname{alph}\xi$. Thus $g(\psi(wx)) = g(\psi(\delta)) = v \in \operatorname{Fact}\Pi$ and $\psi(wx) \in \operatorname{Pref} t$, but this contradicts the maximality of w. Therefore $\zeta \neq \varepsilon$, so that $g(\psi(w'x))$ is left special in s, being preceded both by $b \notin \operatorname{Suff}\Pi$ and by $(g(\zeta))^\ell \in \operatorname{Suff}\Pi$. Hence $g(\psi(w'x))$ is a prefix of s, and then of $g(\psi(w\xi x))$. By Lemma 1.3.2, we obtain $\psi(w'x) \in \operatorname{Pref}\psi(w\xi x)$, a contradiction. Thus ϕ satisfies condition 3 of Theorem 4.3.1.

Finally, let $u = g(\psi(w)) \in \operatorname{Pref} s$ and let us prove that $LS(\{u\} \cup \Pi) \subseteq \operatorname{Pref} u$. Any word $\lambda \in LS(\{u\} \cup \Pi)$ is left special in s, and hence a prefix of it. If λ is a factor of u, then $|\lambda| \leq |u|$, so that $\lambda \in \operatorname{Pref} u$ as desired.

Let then $\lambda \in LS\,\Pi$, with $\lambda \neq \varepsilon$. Since $\lambda \in \operatorname{Pref} s$, we have $\lambda \in \operatorname{Pref}\Pi^*$, so that by Proposition 1.3.5 there exists a unique $\lambda' = \pi_1\pi_2\cdots\pi_k \in \Pi^*$ (with $k \geq 1$ and $\pi_i \in \Pi$ for $i = 1, \ldots, k$) such that $\lambda \in \operatorname{Pref}\lambda'$ and $\pi_1\cdots\pi_{k-1} \in \operatorname{Pref}\lambda$. Because of its uniqueness, λ' has to be a prefix of s. Moreover, as a consequence of Proposition 1.3.1, every occurrence of λ as a factor of any $\pi \in \Pi$ can be extended to the right to $\lambda' \in \operatorname{Fact}\pi$, so that $\lambda' \in LS\,\Pi$. As $\lambda' \in \Pi^*$, we can write $\lambda' = g(\tau) \in \operatorname{Pref} g(t)$ for some $\tau \in B^*$. By Lemma 1.3.2, τ is a prefix of t.

As $\lambda' \in LS\,\Pi$, it is a proper factor of some $\pi \in \Pi$. By Proposition 1.3.6, we can write $\pi = hp\lambda' qh'$ with $h, h' \in A^+$, $p, q \in \Pi^*$, $b = h^\ell \notin \operatorname{Suff}\Pi$, and $\bar{c} = (h')^f \notin \operatorname{Pref}\Pi$. Therefore, as we have already proved that condition 3 of Theorem 4.3.1 is satisfied, $p\lambda' q = g(\psi(w'x))$ for suitable $w' \in \operatorname{Pref} w$ and $x \in \{\varepsilon\} \cup (B \setminus X)$. As $p \in \Pi^*$, this implies $\tau \in \operatorname{Fact}\psi(w'x)$.

We claim that $\tau \in \operatorname{Pref}\psi(w)$, so that $\lambda \in \operatorname{Pref}\lambda'$ is a prefix of u. Indeed, suppose this is not the case, so that, since $\tau \in \operatorname{Pref} t$, one has $\psi(w)d \in \operatorname{Pref}\tau$ where d is the first letter of Δ. Then $\psi(w)d \in \operatorname{Fact}\psi(w'x)$. This is absurd if $w'x \in \operatorname{Pref} w$, as $|\psi(w)d| > |\psi(w'x)|$ in that case. If $w'x \notin \operatorname{Pref} w$, since $w' \in \operatorname{Pref} w$ we can write $w = w'yw''$ for some letter $y \neq x$ and $w'' \in B^*$. Then $\psi(w')y$ is a prefix of $\psi(w)d \in \operatorname{Fact}\psi(w'x) \subseteq \operatorname{Fact}\psi(w'x^\omega)$. As $y \notin \operatorname{alph} x^\omega$, we reach a contradiction by Proposition 2.2.12. Hence all conditions of Theorem 4.3.1 are met, so that ϕ is ϑ-characteristic. $\square$

Let us consider the family $SW_\vartheta(N)$, introduced in [12], of all words $w \in A^\omega$ which are closed under ϑ and such that every left special factor of w whose length is at least N is a prefix of w. Moreover, SW_ϑ will denote the class of words which are in $SW_\vartheta(N)$ for some $N \geq 0$. One has that $SW_\vartheta(0) = SEpi_\vartheta$. It has been proved in [12] that the family of ϑ-standard words is included in $SW_\vartheta(3)$, and that SW_ϑ coincides with the family of *ϑ-standard words with seed* introduced in [27, 13].

Proposition 4.4.2. *Let $\varphi : X^* \to A^*$ be an injective morphism decomposable as $\varphi = g \circ \mu_w \circ \eta$ where $w \in B^*$, $B = \operatorname{alph} w \cup \eta(X)$, η a literal morphism, and g is an injective morphism such that $g(B) = \Pi \subseteq \mathcal{P}_\vartheta$. If Π is overlap-free and normal, then $\varphi(SEpi(X)) \subseteq SW_\vartheta(N)$ with $N = \max\{|\pi| \mid \pi \in \Pi\}$.*

Proof. The proof is very similar to the sufficiency of Theorem 4.3.1. Using the same notation, suppose that λ is a left special factor of $g(t)$ of length $|\lambda| \geq N$

where $t = \mu_w(t') \in SEpi(B)$ and $t' \in SEpi(\eta(X))$. One has that Cases 1 and 3 cannot occur since otherwise one would derive $a\lambda \in \mathrm{Fact}\,\Pi$ that implies $|\lambda| < N$, which is a contradiction. It remains to consider Case 2. By using exactly the same argument one obtains that λ is a prefix of $g(t)$. Finally, since $g(t)$ has infinitely many ϑ-palindromic prefixes one has that $g(t)$ is closed under ϑ. $\square$

Chapter 5

Rich words

In this chapter, we present and study the class of words which have a maximal number of palindromic factors for each length. Such words, called *rich words*, are again a generalization of Episturmian words, though quite different from those presented so far. The main results of this chapter have appeared in [16] and [15].

5.1 Introduction

Given an infinite word $\mathbf{w}$, let $\mathcal{P}(n)$ (resp. $\mathcal{C}(n)$) denote the *palindromic complexity* (resp. *factor complexity*) of $\mathbf{w}$, i.e., the number of distinct palindromic factors (resp. factors) of $\mathbf{w}$ of length n. In [1], J.-P. Allouche, M. Baake, J. Cassaigne, and D. Damanik established the following inequality relating the palindromic and factor complexities of a non-ultimately periodic infinite word:

$$\mathcal{P}(n) \leq \frac{16}{n}\mathcal{C}\left(n + \left\lfloor \frac{n}{4} \right\rfloor\right) \quad \text{for all } n \in \mathbb{N}.$$

More recently, using *Rauzy graphs*, P. Baláži, Z. Masáková, and E. Pelantová [5] proved that for any uniformly recurrent infinite word whose set of factors is closed under reversal,

$$\mathcal{P}(n) + \mathcal{P}(n+1) \leq \mathcal{C}(n+1) - \mathcal{C}(n) + 2 \quad \text{for all } n \in \mathbb{N}. \tag{5.1}$$

They also provided several examples of infinite words for which $\mathcal{P}(n)+\mathcal{P}(n+1)$ always reaches the upper bound given in relation (5.1). Such infinite words include *Arnoux-Rauzy sequences*, *complementation-symmetric sequences*, certain words associated with β-expansions where β is a *simple Parry number*, and a class of words coding r-interval exchange transformations.

In this section we give a characterization of all infinite words with factors closed under reversal for which the equality $\mathcal{P}(n)+\mathcal{P}(n+1) = \mathcal{C}(n+1)-\mathcal{C}(n)+2$ holds for all n: these are exactly the infinite words with the property that all 'complete returns' to palindromes are palindromes. Given a finite or infinite word w and a factor u of w, we say that a factor r of w is a *complete return* to u in w if r contains exactly two occurrences of u, one as a prefix and one as a suffix. Return words play an important role in the study of minimal subshifts; see [31, 32, 33, 34, 40, 54].

Our main theorem is the following:

Theorem 5.1.1. *For any infinite word* $\mathbf{w}$ *whose set of factors is closed under reversal, the following conditions are equivalent:*

(I) *all complete returns to any palindromic factor of* $\mathbf{w}$ *are palindromes;*

(II) $\mathcal{P}(n) + \mathcal{P}(n+1) = \mathcal{C}(n+1) - \mathcal{C}(n) + 2$ *for all* $n \in \mathbb{N}$.

Recently, in [38], it was shown that property (I) is equivalent to every factor u of $\mathbf{w}$ having exactly $|u|+1$ distinct palindromic factors (including the empty word). Such words are 'rich' in palindromes in the sense that they contain the maximum number of different palindromic factors. Indeed, X. Droubay, J. Justin, and G. Pirillo [29] observed that any finite word w of *length* $|w|$ contains at most $|w|+1$ distinct palindromes.

As already said, the family of finite and infinite words having property (I) are called *rich words* in [38]. In independent work, P. Ambrož, C. Frougny, Z. Masáková, and E. Pelantová [2] have considered the same class of words which they call *full words*, following earlier work of S. Brlek, S. Hamel, M. Nivat, and C. Reutenauer in [10].

Rich words encompass the well-known family of *Episturmian words* originally introduced by X. Droubay, J. Justin, and G. Pirillo in [29] (see Section 5.3 for more details). Another special class of rich words consists of S. Fischler's sequences with "abundant palindromic prefixes", which were introduced and studied in [35] in relation to Diophantine approximation (see also [36]). Other examples of rich words that are neither Episturmian nor of "Fischler type" include: non-recurrent rich words, like $abbbb\cdots$ and $abaabaaabaaaab\cdots$; the periodic rich infinite words: $(aab^kaabab)(aab^kaabab)\cdots$, with $k \geq 0$; the non-ultimately periodic recurrent rich infinite word $\psi(\mathbf{f})$ where $\mathbf{f} = abaababaaba\cdots$ is the *Fibonacci word* and ψ is the morphism: $a \mapsto aab^kaabab$, $b \mapsto bab$; and the recurrent, but not uniformly recurrent, rich infinite word generated by the morphism: $a \mapsto aba$, $b \mapsto bb$. (See [38] for these examples and more.)

From the work in [29, 38], we have the following equivalences.

Proposition 5.1.2. *A finite or infinite word w is* rich *if equivalently:*

- *all complete returns to any palindromic factor of* $\mathbf{w}$ *are palindromes;*
- *every factor u of w contains $|u|+1$ distinct palindomes;*
- *the longest palindromic suffix of any prefix p of w occurs exactly once in p.*

From the perspective of richness, our main theorem can be viewed as a characterization of *recurrent rich infinite words* since any rich infinite word is recurrent if and only if its set of factors is closed under reversal (see [38] or Remark 5.1.1). Interestingly, the proof of Theorem 5.1.1 relies upon another new characterization of rich words (Proposition 5.1.4), which is useful for establishing the key step, namely that the so-called *super reduced Rauzy graph* is a tree. This answers a claim made in the last few lines of [5] where it was remarked that the Rauzy graphs of words satisfying equality (II) must have a very special form.

After some preliminary definitions and results in the next section, Section 3 is devoted to the proof of Theorem 5.1.1 and some interesting consequences are proved in Section 4.

5.1.1 Notation and terminology

We recall that a factor of an infinite word $\mathbf{w}$ is *recurrent* in $\mathbf{w}$ if it occurs infinitely often in $\mathbf{w}$, and $\mathbf{w}$ itself is said to be *recurrent* if all of its factors are recurrent in it. Furthermore, $\mathbf{w}$ is *uniformly recurrent* if any factor of $\mathbf{w}$ occurs infinitely many times in $\mathbf{w}$ with bounded gaps.

Remark. A noteworthy fact (proved in [38]) is that a rich infinite word is recurrent if and only if its set of factors is closed under reversal.

More generally, we have the following well-known result:

Proposition 5.1.3 (folklore). *If* w *is an infinite word with* $F(\mathrm{w})$ *closed under reversal, then* w *is recurrent.*

Proof. Consider some occurrence of a factor u in w and let v be a prefix of w containing u. As $F(\mathrm{w})$ is closed under reversal, $\tilde{v} \in F(\mathrm{w})$. Thus, if v is long enough, there is an occurrence of $\tilde{u}$ strictly on the right of this particular occurrence of u in w. Similarly u occurs on the right of this $\tilde{u}$ and thus u is recurrent in w. □

5.1.2 Key results

We now prove two useful results, the first being a new characterization of rich words.

Proposition 5.1.4. *A finite or infinite word w is rich if and only if, for each factor $v \in F(w)$, any factor of w beginning with v and ending with $\tilde{v}$ and not containing v or $\tilde{v}$ as an interior factor is a palindrome.*

Proof. ONLY IF: Consider any factor $v \in F(w)$ and let u be a factor of w beginning with v and ending with $\tilde{v}$ and not containing v or $\tilde{v}$ as an interior factor. If v is a palindrome, then either $u = v = \tilde{v}$ (in which case u is clearly a palindrome), or u is a complete return to v in w, and hence u is (again) a palindrome by Proposition 5.1.2. Now assume that v is not a palindrome.

Suppose by way of contradiction that u is not a palindrome and let p be the longest palindromic suffix of u (which is unioccurrent in u by richness). Then $|p| < |u|$ as u is not a palindrome. If $|p| > |v|$, then $\tilde{v}$ is a proper suffix of p, and hence v is a proper prefix of p. But then v is an interior factor of u, a contradiction. On the other hand, if $|p| \leq |v|$, then $|p| \neq |v|$ and p is a proper suffix of $\tilde{v}$ (as $\tilde{v}$ is not a palindrome), and hence p is a proper prefix of v. Thus p is both a prefix and a suffix of u; in particular p is not unioccurrent in u, a contradiction.

IF: The given conditions tell us that any complete return to a palindromic factor v $(= \tilde{v})$ of w is a palindrome. Hence w is rich by Proposition 5.1.2. □

Proposition 5.1.5. *Suppose w is a rich word. Then, for any non-palindromic factor v of w, $\tilde{v}$ is a unioccurrent factor of any complete return to v in w.*

Proof. Let r be a complete return to v in w and let p be the longest palindromic suffix of r. Then $|p| > |v|$; otherwise, if $|p| \leq |v|$, then p would occur at least twice in r (as a suffix of each of the two occurrences of v in r), which is impossible as r is rich. Thus v is a proper suffix of p, and hence $\tilde{v}$ is a proper prefix of p. So $\tilde{v}$ is clearly an interior factor of r.

It remains to show that $\tilde{v}$ is unioccurrent in r. Arguing by contradiction, we suppose that $\tilde{v}$ occurs more than once in r. Then a complete return r' to $\tilde{v}$ occurs as a proper factor of r. Using the same reasoning as above, v is an interior factor of r', and hence an interior factor of r, contradicting the fact that r is a complete return to v. Thus $\tilde{v}$ is unioccurent in r. $\square$

Note. The above proposition tells us that for any factor v of a rich word w, occurrences of v and $\tilde{v}$ alternate in w.

5.1.3 Proof of Theorem 5.1.1

Following the method of Baláži *et al.* [5], a key tool for the proof of our main theorem is the notion of a *Rauzy graph*, defined as follows. Given an infinite word w, the *Rauzy graph of order n* for w, denoted by $\Gamma_n(\mathrm{w})$, is the directed graph with set of vertices $F_n(\mathrm{w})$ and set of edges $F_{n+1}(\mathrm{w})$ such that an edge $e \in F_{n+1}(\mathrm{w})$ starts at vertex v and ends at a vertex v' if and only if v is a prefix of e and v' is a suffix of e. For a vertex v, the *out-degree* of v (denoted by $\deg^+(v)$) is the number of distinct edges leaving v, and the *in-degree* of v (denoted by $\deg^-(v)$) is the number of distinct edges entering v. More precisely:

$$\deg^+(v) = \sharp\{x \in \mathcal{A} \mid vx \in F_{n+1}(\mathrm{w})\} \quad \text{and} \quad \deg^-(v) = \sharp\{x \in \mathcal{A} \mid xv \in F_{n+1}(\mathrm{w})\}.$$

We observe that, for all $n \in \mathbb{N}$,

$$\sum_{v \in F_n(\mathrm{w})} \deg^+(v) = \sharp F_{n+1}(\mathrm{w}) = \sum_{v \in F_n(\mathrm{w})} \deg^-(v).$$

(Note that $\sharp F_{n+1}(\mathrm{w}) = \mathcal{C}(n+1)$.) Hence

$$\mathcal{C}(n+1) - \mathcal{C}(n) = \sum_{v \in F_n(\mathrm{w})} (\deg^+(v) - 1) = \sum_{v \in F_n(\mathrm{w})} (\deg^-(v) - 1). \tag{5.2}$$

It is therefore easy to see that a factor $v \in F_n(\mathbf{w})$ positively contributes to $\mathcal{C}(n+1) - \mathcal{C}(n)$ if and only if $\deg^+(v) \geq 2$, i.e., if and only if there exist at least two distinct letters a, b such that va, $vb \in F_{n+1}(\mathbf{w})$, in which case v is said to be a *right-special* factor of $\mathbf{w}$. Similarly, a factor $v \in F_n(\mathbf{w})$ is said to be a *left-special* factor of $\mathbf{w}$ if there exist at least two distinct letters a, b such that av, $bv \in F_{n+1}(\mathbf{w})$. A factor of $\mathbf{w}$ is said to be *special* if it is either left-special or right-special (not necessarily both). With this terminology, if we let $S_n(\mathbf{w})$ denote the set of special factors of $\mathbf{w}$ of length n, then formula (5.2) may be expressed as:

$$\mathcal{C}(n+1) - \mathcal{C}(n) = \sum_{v \in S_n(\mathbf{w})} (\deg^+(v) - 1) \quad \text{for all } n \in \mathbb{N}. \tag{5.3}$$

Using similar terminology to that in [5], a directed path P in the Rauzy graph $\Gamma_n(\mathbf{w})$ is said to be a *simple path of order* n if it begins with a special factor v and ends with a special factor v' and contains no other special factors, i.e., P is a directed path of the form vv' or $vz_1 \cdots z_k v'$ where each z_i is a non-special factor of length n. A special factor $v \in S_n(\mathbf{w})$ is called a *trivial* simple path of order n.

In what follows, we use the following terminology for paths. Hereafter, "path" should be taken to mean "directed path".

Definition 5.1.6. *Suppose* $\mathbf{w}$ *is an infinite word and let* $P = v \cdots v'$ *be a path in* $\Gamma_n(\mathbf{w})$.

- *The first vertex* v *(resp. last vertex* v'*) is called the* initial vertex *(resp.* terminal vertex*) of* P.
- *A vertex of* P *that is neither an initial vertex nor a terminal vertex of* P *is called an* interior vertex *of* P.
- P *is said to be a* non-trivial path *if it consists of at least two distinct vertices.*
- *The* reversal $\tilde{P}$ *of the path* P *is the path obtained from* P *be reversing all edge labels (and arrows) and all labels of vertices.*
- *We say that* P *is* palindromic *(or that* P *is* invariant under reversal*) if* $P = \tilde{P}$.

Note. Given a path P in $\Gamma_n(\mathbf{w})$, the reversal of P does not necessarily exist in $\Gamma_n(\mathbf{w})$.

Suppose $P = w_1w_2\cdots w_k$ is a non-trivial path in $\Gamma_n(\mathbf{w})$, and for each i with $1 \leq i \leq k$, let a_i and b_i denote the respective first and last letters of w_i. Then, by the definition of $\Gamma_n(\mathbf{w})$, we have $w_1b_2\cdots b_k = a_1\cdots a_{k-1}w_k$. We call this word the *label* of the path P, denoted by ℓ_P. Note that the i-th shift of $\ell_P := w_1b_2\cdots b_k$ begins with w_{i+1} for all i with $1 \leq i \leq k-1$.

For our purposes, it is convenient to consider the *reduced Rauzy graph of order n*, denoted by $\Gamma'_n(\mathbf{w})$, which is the directed graph obtained from $\Gamma_n(\mathbf{w})$ by replacing each *simple path* $P = w_1w_2\cdots w_{k-1}w_k$ with a directed edge $w_1 \to w_k$ labelled by ℓ_P. Thus the set of vertices of $\Gamma'_n(\mathbf{w})$ is $S_n(\mathbf{w})$. For example, consider the (rich) *Fibonacci word*:

$$\mathbf{f} = abaababaabaababaababaabaababaabaababaababaabaababaaba\cdots$$

which is generated by the *Fibonacci morphism* $\varphi : a \mapsto ab, b \mapsto a$. The reduced Rauzy graph $\Gamma'_2(\mathbf{f})$ consists of the two (special) vertices: ab, ba and three directed edges: $ab \to ba$, $ba \to ba$, $ba \to ab$ with respective labels: aba, $baab$, bab.

Lemma 5.1.7. *Let $\mathbf{w}$ be a rich infinite word and suppose $P = w_1w_2\cdots w_k$ is a non-trivial path in $\Gamma_n(\mathbf{w})$ with $k \geq 2$. Then the label $\ell_P = w_1b_2\cdots b_k$ is a rich word.*

Proof. We proceed by induction on the number of vertices k in P. The lemma is clearly true for $k = 2$ since $\ell_P = w_1b_2$ is a factor of $\mathbf{w}$ of length $n+1$. Now suppose $k \geq 3$ and assume that the label of any path consisting of $k-1$ vertices is rich. Consider any path consisting of k vertices, namely $P = w_1w_2\cdots w_k$, and suppose by way of contradiction that its label $\ell_P = w_1b_2\cdots b_k$ is not rich. Then the longest palindromic prefix p of ℓ_P occurs more than once in ℓ_P. Hence there exists a complete return r to p which is a prefix of ℓ_P. It follows that $r = \ell_P$, otherwise r would be a factor of the prefix $u := w_1b_2\cdots b_{k-1}$ of ℓ_P, and hence a palindrome since u is rich by the induction hypothesis. But this contradicts the maximality of the palindromic prefix p. So ℓ_P is a non-palindromic complete return to p. Let q be the longest palindromic prefix of u (which is unioccurrent in u by richness). If $|p| > |q|$, then q is a proper prefix of p, and hence q occurs more than twice in u, a contradiction. On the other hand, if $|p| \leq |q|$, then p is a prefix of q, and hence p is an interior factor of ℓ_P (occurring as a suffix of q), a contradiction. Thus ℓ_P is rich, as required. $\square$

The proof of Theorem 5.1.1 relies upon the following extensions of Propositions 5.1.4–5.1.5 to paths.

Lemma 5.1.8. (Analogue of Proposition 5.1.4.) *Suppose* $\mathbf{w}$ *is a rich infinite word and let* v *be any factor of* $\mathbf{w}$ *of length* n*. If* $P = v \cdots \tilde{v}$ *is a path from* v *to* $\tilde{v}$ *in* $\Gamma_n(\mathbf{w})$ *that does not contain* v *or* $\tilde{v}$ *as an interior vertex, then* P *is palindromic. This property also holds for paths in* $\Gamma'_n(\mathbf{w})$*.*

Proof. We first observe that if P consists of a single vertex, then $P = v = \tilde{v}$, and hence P is palindromic. Now suppose P is a non-trivial path. If $P = v\tilde{v}$, then P is clearly palindromic. So suppose $P = vz_1 \cdots z_k\tilde{v}$ where the z_i are factors of $\mathbf{w}$ of length n. By definition, the label $\ell_P = vb_1 \cdots b_k b_{k+1}$ begins with v and ends with $\tilde{v}$ and contains neither v nor $\tilde{v}$ as an interior factor (otherwise P would contain v or $\tilde{v}$ as an interior vertex, which is not possible). Thus, as ℓ_P is rich (by Lemma 5.1.7), it follows that ℓ_P is a palindrome by Proposition 5.1.4; whence P must be invariant under reversal too. It is easy to see that this property is also true for paths in the reduced Rauzy graph $\Gamma'_n(\mathbf{w})$. $\square$

Lemma 5.1.9. (Analogue of Proposition 5.1.5.) *Suppose* $\mathbf{w}$ *is a rich infinite word and let* v *be any non-palindromic factor of* $\mathbf{w}$ *of length* n*. If* $P = v \cdots v$ *is a non-trivial path in* $\Gamma_n(\mathbf{w})$ *that does not contain* v *as an interior vertex, then* P *passes through* $\tilde{v}$ *exactly once. This property also holds for paths in* $\Gamma'_n(\mathbf{w})$*.*

Note. Of particular usefulness is the fact that any path from v to v must pass through $\tilde{v}$.

Proof. Let us write $P = vz_1 \cdots z_k v$ where the z_i are factors of $\mathbf{w}$ of length n. By definition, the label $\ell_P = vb_1 \cdots b_k b_{k+1}$ contains exactly two occurrences of v, one as a prefix and one as a suffix (otherwise, if ℓ_P contained v as an interior factor, then v would be an interior vertex of P, which is not possible). Thus, as ℓ_P is rich (by Lemma 5.1.7), it follows that $\tilde{v}$ is a unioccurrent (interior) factor of ℓ_P by Proposition 5.1.5; whence P passes through $\tilde{v}$ exactly once. It is easy to see that this property is also true for paths in the reduced Rauzy graph $\Gamma'_n(\mathbf{w})$. $\square$

(I) implies (II)

Suppose $\mathbf{w}$ is an infinite word with $F(\mathbf{w})$ closed under reversal and satisfying property (I). Then $\mathbf{w}$ is recurrent by Proposition 5.1.3 (i.e., $\mathbf{w}$ is a recurrent rich infinite word). Moreover, recurrence implies that for all n, the Rauzy graph $\Gamma_n(\mathbf{w})$ is strongly connected, i.e., there exists a directed path from any vertex v to every other vertex v' in $\Gamma_n(\mathbf{w})$.

Fix $n \in \mathbb{N}$ and let us now consider the *super reduced Rauzy graph of order* n, denoted by $\Gamma''_n(\mathbf{w})$, whose set of vertices consists of all $[v] := \{v, \tilde{v}\}$ where v is any special factor of length n. Any two distinct vertices $[v]$, $[w]$ (with $v \notin \{w, \tilde{w}\}$) are joined by an undirected edge with label $[\ell_P] := \{\ell_P, \ell_{\tilde{P}}\}$ if P or $\tilde{P}$ is a simple path beginning with v or $\tilde{v}$ and ending with w or $\tilde{w}$. For example, in the case of the Fibonacci word, $\Gamma''_2(\mathbf{f})$ consists of only one vertex: $[ab]$. In general, the super reduced Rauzy graph consists of more than one vertex and may contain multiple edges between vertices.

Suppose $\Gamma''_n(\mathbf{w})$ consists of s vertices; namely $[v_i]$, $i = 1, \ldots, s$. Since $\Gamma_n(\mathbf{w})$ is strongly connected (by recurrence), $\Gamma''_n(\mathbf{w})$ is connected; thus it contains at least $s - 1$ edges.

Now, from Lemma 5.1.8, we know that if v is a special factor, any simple path from v to $\tilde{v}$ is palindromic (i.e., invariant under reversal). Moreover, by closure under reversal, if there exists a simple path P from a special factor v to a special factor w, with $v \notin \{w, \tilde{w}\}$, then there is also a simple path from $\tilde{w}$ to $\tilde{v}$ (namely, the reversal of the path P). Neither of these simple paths is palindromic.

We thus deduce that there exist *at least* $2(s-1)$ non-trivial simple paths in the Rauzy graph $\Gamma_n(\mathbf{w})$ that are non-palindromic (i.e., not invariant under reversal). In fact, we will show that there are exactly $2(s-1)$ non-trivial simple paths of order n that are non-palindromic. Indeed, if this true then, as each palindromic factor of length n or $n+1$ is a central factor of a (unique) palindromic simple path of order n, we have:

$$\mathcal{P}(n) + \mathcal{P}(n+1) = \sum_{v \in S_n(\mathbf{w})} \deg^+(v) - 2(s-1) + p \tag{5.4}$$

where, on the right hand side, the first summand is the total number of non-trivial simple paths, the second summand is the number of non-trivial simple paths that are non-palindromic, and p is the number of special palindromes of length n (i.e., the number of trivial simple paths of order n that are palindromic). By observing that the number of special factors of length n is $2s-p$, we can simplify equation (5.4) to obtain the required equality (II) as follows:

$$\begin{aligned} \mathcal{P}(n) + \mathcal{P}(n+1) &= \sum_{v \in S_n(\mathbf{w})} \deg^+(v) - (2s-p) + 2 \\ &= \sum_{v \in S_n(\mathbf{w})} (\deg^+(v) - 1) + 2 \\ &= \mathcal{C}(n+1) - \mathcal{C}(n) + 2 \qquad \text{(by (5.3))}. \end{aligned}$$

We observe, in particular, that any infinite word $\mathbf{w}$ with $F(\mathbf{w})$ closed under reversal satisfies equality (II) if and only if any simple path between a special factor and its reversal is palindromic, and for each n, there are exactly $2(s-1)$ non-trivial simple paths of order n that are non-palindromic. The latter condition says that, for all n, the super reduced Rauzy graph $\Gamma''_n(\mathbf{w})$ contains exactly $s-1$ edges (with each edge corresponding to a simple path and its reversal), and hence $\Gamma''_n(\mathbf{w})$ is a tree as it contains s vertices, $s-1$ edges, and must be connected by the recurrence of $\mathbf{w}$ (which follows from Proposition 5.1.3). More formally:

Proposition 5.1.10. *An infinite word* $\mathbf{w}$ *with* $F(\mathbf{w})$ *closed under reversal satisfies equality* (II) *if and only if the following conditions hold:*

1) *any simple path between a special factor and its reversal is palindromic;*

2) *the super reduced Rauzy graph* $\Gamma''_n(\mathbf{w})$ *is a tree for all* n.

Proof. Suppose $\mathbf{w}$ is an infinite word with $F(\mathbf{w})$ closed under reversal. Then $\mathbf{w}$ is recurrent by Proposition 5.1.3. We have already shown that conditions 1) and 2) imply that $\mathbf{w}$ satisfies equality (II). Conversely, if at least one of conditions 1) and 2) does not hold, then $\mathcal{P}(n) + \mathcal{P}(n+1) < \mathcal{C}(n+1) - \mathcal{C}(n) + 2$ (by the arguments preceding this proposition), i.e., $\mathbf{w}$ does not satisfies equality (II). $\square$

To complete the proof of "(I) $\Rightarrow$ (II)", it remains to show that any recurrent rich infinite word $\mathbf{w}$ satisfies condition 2) of Proposition 5.1.10, since we have already shown that condition 1) holds for any such $\mathbf{w}$ (using Lemma 5.1.8). The proof of the fact that $\mathbf{w}$ satisfies condition 2) uses the following two lemmas (Lemmas 5.1.11–5.1.12).

Notation. Given two distinct special factors v, w of the same length n, we write $v \not\to w$ if there does not exist a directed edge from v to w in the reduced Rauzy graph $\Gamma'_n(\mathbf{w})$ (i.e., if there does not exist a simple path from v to w).

Lemma 5.1.11. *Suppose $\mathbf{w}$ is a recurrent rich infinite word and let v, w be two distinct special factors of $\mathbf{w}$ of the same length with $v \notin \{w, \tilde{w}\}$. If there exists a simple path P from v to w, then P is unique and there also exists a unique simple path from $\tilde{w}$ to $\tilde{v}$ (namely, the reversal of P). Moreover:*

i) $v \not\to \tilde{w}$, and hence $w \not\to \tilde{v}$ (unless w is a palindrome);

ii) $\tilde{w} \not\to v$, and hence $\tilde{v} \not\to w$ (unless v is a palindrome);

iii) $w \not\to v$, and hence $\tilde{v} \not\to \tilde{w}$ (unless v and w are both palindromes).

Proof. By closure under reversal (Remark 5.1.1), if there exists a simple path P from v to w, then the reversal of P is a simple path from $\tilde{w}$ to $\tilde{v}$ in the Rauzy graph of order $|v| = |w| = n$. To prove the uniqueness of P, let us suppose there exist two different simple paths P_1, P_2 from v to w in the Rauzy graph $\Gamma_n(\mathbf{w})$. Then

$$P_1 = vu_1 \cdots u_k w \quad \text{and} \quad P_2 = vz_1 \cdots z_\ell w \quad \text{for some } k,\, l \in \mathbb{N},$$

where $u_1, \ldots, u_k, z_1, \ldots, z_\ell$ are non-special factors of $\mathbf{w}$ of length n and $u_i \neq z_i$ for some i. Note that either P_1 or P_2 (not both) may be of the form vw.

To keep the rest of the proof as simple as possible, we assume hereafter that neither v nor w is a palindrome; the arguments are similar, and in fact easier, in the cases when either v or w (or both) is a palindrome.

Consider a path Q of minimal length beginning with P_1 and ending with P_2 (in the Rauzy graph $\Gamma_n(\mathbf{w})$):

$$Q = P_1 \cdots P_2 = vu_1 \cdots u_k \underbrace{w \cdots v}_{Q_1} z_1 \cdots z_\ell w.$$

First we observe that Q contains $\tilde{v}$ since any path from v to itself must pass through $\tilde{v}$, by Lemma 5.1.9. Moreover, the left-most $\tilde{v}$ in Q must occur in the subpath Q_1 (since $\tilde{v}$ is not equal to any of the non-special factors u_i, z_j and $\tilde{v} \neq w$). Therefore

$$Q = \underbrace{vu_1 \cdots u_k w \cdots \tilde{v}}_{Q_2} \cdots vz_1 \cdots z_\ell w$$

where the subpath Q_2 ends with the left-most $\tilde{v}$ in the path Q. By Lemma 5.1.9, Q_2 is a path from v to $\tilde{v}$ that does not contain v or $\tilde{v}$ as an interior vertex. Thus, by Lemma 5.1.8, Q_2 is palindromic, and hence Q_2 ends with the reversal of the path P_1 since it begins with P_1. More explicitly:

$$Q = \underbrace{vu_1 \cdots u_k w}_{P_1} \cdots \overbrace{\underbrace{\tilde{w}\tilde{u}_k \cdots \tilde{u}_1 \tilde{v}}_{\widetilde{P}_1} \cdots \underbrace{vz_1 \cdots z_\ell w}_{P_2}}^{Q_3}.$$

We distinguish two cases.

Case 1: If the subpath Q_3 contains w as a terminal vertex only, then $\tilde{w}$ is not an interior vertex of Q_3 by Lemma 5.1.9, and hence Q_3 is palindromic by Lemma 5.1.8. It follows that $k = \ell$ and $z_i = u_i$ for all $i = 1, \ldots, k$. Thus $P_1 = P_2$; a contradiction.

Case 2: If the subpath Q_3 contains w as an interior vertex, then Q_3 first passes through w after taking the path $\widetilde{P}_1$ (at the beginning) and before taking the path P_2 (at the end). Hence, by Lemma 5.1.8, Q_3 begins with a palindromic path from $\tilde{w}$ to w that begins with $\widetilde{P}_1$ and hence ends with P_1. But then Q passes through the path P_1 at least twice before taking the path P_2, contradicting the fact that Q is a path of minimal length beginning with P_1 and ending with P_2.

Both cases lead to a contradiction; thus the simple path P from v to w is unique (and its reversal $\widetilde{P}$ is the unique simple path from $\tilde{w}$ to $\tilde{v}$). It remains to show that conditions *i*)–*iii*) hold. As *ii*) is symmetric to *i*), we prove only that i) and iii) are satisfied. By what precedes, it suffices to consider paths in the reduced Rauzy graph $\Gamma'_n(\mathbf{w})$.

i): Arguing by contradiction, let us suppose that there exists a (unique) simple path from v to $\tilde{w}$, i.e., there exists a directed edge from v to $\tilde{w}$ in the reduced Rauzy graph $\Gamma'_n(\mathbf{w})$. Then (from above) we know that there also exists a directed edge from w to $\tilde{v}$. Consider a shortest path Q in the reduced Rauzy graph $\Gamma'_n(\mathbf{w})$ beginning with $v\tilde{w}$ and ending with vw. By Lemma 5.1.9, any path from v to itself passes through $\tilde{v}$, so we may write

$$Q = \underbrace{v\tilde{w}\cdots\tilde{v}}_{Q_1}\cdots vw,$$

where the subpath Q_1 ends with the left-most $\tilde{v}$ in the path Q. By Lemmas 5.1.8–5.1.9, the path $Q_1 = v\tilde{w}\cdots\tilde{v}$ is palindromic, and hence it ends with $w\tilde{v}$. So we have $Q = v\tilde{w}\cdots w\tilde{v}\cdots vw$; moreover, by Lemma 5.1.9, $\tilde{w}$ must occur between the last two w's shown here. In particular,

$$Q = v\tilde{w}\cdots\underbrace{w\tilde{v}\cdots\tilde{w}}_{Q_2}\cdots vw$$

where the subpath Q_2 contains $\tilde{w}$ as a terminal vertex only. Thus, by Lemmas 5.1.8–5.1.9, the path $Q_2 = w\tilde{v}\cdots\tilde{w}$ is palindromic, and hence it ends with $v\tilde{w}$. But then Q ends with a shorter path of the form $v\tilde{w}\cdots vw$, contradicting the fact that Q is a path of minimal length beginning with $v\tilde{w}$ and ending with vw.

iii): Again, the proof proceeds by contradiction. Suppose there exists a (unique) simple path from w to v. Consider a shortest path Z in the reduced Rauzy graph $\Gamma'_n(\mathbf{w})$ beginning with wv and ending with vw. By Lemma 5.1.9, the path Z must pass through $\tilde{w}$; thus

$$Z = \underbrace{wv\cdots\tilde{w}}_{Z_1}\cdots vw.$$

where the subpath Z_1 ends with the left-most $\tilde{w}$ in the path Z. Now it follows from Lemmas 5.1.8–5.1.9 that the subpath Z_1 is palindromic, and hence Z_1 must end with $\tilde{v}\tilde{w}$. So we may write

$$Z = wv\cdots\underbrace{\tilde{v}\tilde{w}\cdots v}_{Z_2}w.$$

If the subpath Z_2 contains v as a terminal vertex only, then neither v nor $\tilde{v}$ is an interior vertex of Z_2 by Lemma 5.1.9. Thus Z_2 is palindromic by Lemma 5.1.8, and hence Z_2 ends with wv. But then the path Z ends with the path wvw, which is impossible by Lemma 5.1.9. Thus, the subpath Z_2 must pass through v at an earlier point, and hence we have $Z_2 = \tilde{v}\tilde{w}\cdots v\cdots v$. In particular, the path Z_2 begins with a palindromic subpath of the form $\tilde{v}\tilde{w}\cdots wv$, by Lemma 5.1.8. But then the path Z ends with a shorter path from wv to vw, contradicting the minimality of Z. $\square$

Notation. For a finite word v, let v^ϵ represent either v or $\tilde{v}$ and set $v^{-\epsilon} := \widetilde{v^\epsilon}$.

Lemma 5.1.12. *Let* $\mathbf{w}$ *be a recurrent rich infinite word. For fixed* $n \in \mathbb{N}_+$*, suppose the super reduced Rauzy graph* $\Gamma''_n(\mathbf{w})$ *contains at least three distinct vertices:* $[v_1]$, $[v_2]$, $\ldots$, $[v_s]$, $s \geq 3$. *Then, for each* k *with* $3 \leq k \leq s$, *the reduced Rauzy graph* $\Gamma'_n(\mathbf{w})$ *contains a path from* v_1 *to* $v_k^{\epsilon_k}$ *of the form:*

$$v_1 v_2^{\epsilon_2} \cdots v_2 v_3^{\epsilon_3} \cdots v_{k-2} v_{k-1}^{\epsilon_{k-1}} \cdots v_{k-1} v_k^{\epsilon_k},$$

where for all $i = 2, \ldots, k-1$, *the subpath* $v_i^{\epsilon_i} \cdots v_i$ *(which may consist of only the single vertex* $v_i^{\epsilon_i}$*) does not contain* v_j, $\tilde{v}_j$ *for all* j *with* $1 \leq j \leq k$, $j \neq i$.

Proof. We use induction on k and employ similar reasoning to the proof of Lemma 5.1.11.

First consider the case $k = 3$. Recurrence implies that $\Gamma'_n(\mathbf{w})$ is connected, so we may assume without loss of generality that $\Gamma'_n(\mathbf{w})$ contains a directed edge from v_1 to $v_2^{\epsilon_2}$, a directed edge from v_2 to $v_3^{\epsilon_3}$, and a path from $v_2^{\epsilon_2}$ to v_2. That is, $\Gamma'_n(\mathbf{w})$ contains a path beginning with $v_1 v_2^{\epsilon_2}$ and ending with $v_2 v_3^{\epsilon_3}$. Consider such a path of minimal length:

$$Q = v_1 v_2^{\epsilon_2} \cdots v_2 v_3^{\epsilon_3}.$$

To prove the claim for $k = 3$, we show that none of the special factors v_1, $\tilde{v}_1$, v_3, $\tilde{v}_3$ are interior vertices of Q. If $v_2^{\epsilon_2} = v_2$, then $Q = v_1 v_2 v_3^{\epsilon_3}$ (by minimality) and we are done. So let us assume that $v_2^{\epsilon_2} = \tilde{v}_2 \neq v_2$.

Observe that if v_1 is an interior vertex of Q, then $\tilde{v}_1$ must be an interior vertex of Q since any path from v_1 to itself must contain $\tilde{v}_1$, by Lemma 5.1.9.

Similarly, if $v_3^{\epsilon_3}$ is an interior vertex of Q, then $v_3^{-\epsilon_3}$ is an interior vertex of Q. Therefore it suffices to show that $\tilde{v}_1$ and $v_3^{-\epsilon_3}$ are not interior vertices of Q. We prove this fact only for $\tilde{v}_1$ as the proof is similar for $v_3^{-\epsilon_3}$.

Arguing by contradiction, suppose $\tilde{v}_1$ is an interior vertex of Q. Then Q begins with a palindromic path from $v_1\tilde{v}_2$ to $\tilde{v}_1$ (by Lemmas 5.1.8–5.1.9), and this palindromic path clearly ends with $v_2\tilde{v}_1$. Hence

$$Q = v_1\tilde{v}_2 \cdots \underbrace{v_2\tilde{v}_1 \cdots v_2}_{Q'} v_3^{\epsilon_3}$$

where the subpath Q' begins with a palindromic path from $v_2\tilde{v}_1$ to $\tilde{v}_2$ (by Lemmas 5.1.8–5.1.9), and this palindromic path clearly ends with $v_1\tilde{v}_2$. But then the path Q ends with a shorter path from $v_1\tilde{v}_2$ to $v_2v_3^{\epsilon_3}$, contradicting the minimality of Q. Thus the lemma holds for $k = 3$.

Now suppose $4 \leq k \leq s$ and assume the claim holds for $k-1$. Since $\Gamma_n'(\mathrm{w})$ is connected, it contains a path beginning with $v_1v_2^{\epsilon_2} \cdots v_2v_3^{\epsilon_3} \cdots v_{k-2}v_{k-1}^{\epsilon_{k-1}}$ and ending with $v_{k-1}v_k^{\epsilon_k}$ (where the former path satisfies the conditions of the lemma). Consider such a path of minimal length:

$$Z = \underbrace{v_1v_2^{\epsilon_2} \cdots v_2v_3^{\epsilon_3} \cdots v_{k-2}}_{Z_1} \underbrace{v_{k-1}^{\epsilon_{k-1}} \cdots v_{k-1}}_{Z_2} v_k^{\epsilon_k} \tag{5.5}$$

where for all $i = 2, \ldots, k-2$, the subpath $v_i^{\epsilon_i} \cdots v_i$ (which may consist of only the single vertex $v_i^{\epsilon_i}$) does not contain v_j, $\tilde{v}_j$ for all j with $1 \leq j \leq k-1$, $j \neq i$. To prove the induction step, we show that the path Z satisfies the following two conditions:

i) the subpath Z_1 contains neither v_k nor $\tilde{v}_k$;

ii) the subpath $Z_2 = v_{k-1}^{\epsilon_{k-1}} \cdots v_{k-1}$ does not contain v_j, $\tilde{v}_j$ for all j with $1 \leq j \leq k$, $j \neq k-1$.

First suppose that condition *i*) is not satisfied, i.e., Z_1 contains v_k or $\tilde{v}_k$. Without loss of generality we assume that v_k is the right-most of the vertices v_k, $\tilde{v}_k$ appearing in Z_1.

Case 1*:* Suppose $v_k^{\epsilon_k} = v_k \neq \tilde{v}_k$. Then Z ends with a path from v_k to itself, which must pass through $\tilde{v}_k$ by Lemma 5.1.9; moreover, $\tilde{v}_k$ must be an interior vertex of Z_2 (by the choice of v_k). Thus, by Lemmas 5.1.8–5.1.9, Z_2v_k (and hence Z) ends with a palindromic path from $\tilde{v}_k$ to $v_{k-1}v_k$. Hence Z_2 contains $\tilde{v}_k\tilde{v}_{k-1}$, and we have:

$$Z_2 v_k = \underbrace{v_{k-1}^{\epsilon_{k-1}} \cdots \tilde{v}_k \tilde{v}_{k-1}}_{Z_3} \cdots v_{k-1} v_k$$

where the subpath Z_3 ends with a palindromic path from v_{k-1} to $\tilde{v}_k\tilde{v}_{k-1}$ (by Lemmas 5.1.8–5.1.9); thus Z_3 contains $v_{k-1}v_k$. But then Z begins with a shorter path from Z_1 to $v_{k-1}v_k^{\epsilon_k}$, contradicting the minimality of Z.

Case 2*:* Suppose $v_k^{\epsilon_k} = \tilde{v}_k$. Then the path Z $(= Z_1 Z_2 \tilde{v}_k)$ ends with a path of the form:

$$Z_4 = v_k \underbrace{\cdots\cdots}_{\text{no } v_k, \tilde{v}_k} Z_2 \tilde{v}_k.$$

If v_k or $\tilde{v}_k$ is an interior vertex of Z_2, then we reach a contradiction using the same arguments as in Case 1. On the other hand, if neither v_k nor $\tilde{v}_k$ is an interior vertex of Z_2, then Z_4 is palindromic by Lemma 5.1.8. So the path Z_4 begins with $v_k\tilde{v}_{k-1}$ since it ends with $v_{k-1}\tilde{v}_k$. But then $\tilde{v}_{k-1}$ is an interior vertex of Z_1, a contradiction.

Thus the path Z satisfies condition i). In proving this fact, we have also shown that v_k, $\tilde{v}_k$ are not interior vertices of Z_2. It remains to show that the subpath Z_2 does not contain v_j, $\tilde{v}_j$ for all j with $1 \leq j \leq k-2$ (and hence Z satisfies condition ii)). We prove only that Z_2 does not contain $\tilde{v}_1$ or $\tilde{v}_1$ since the proof is similar when considering other v_j, $\tilde{v}_j$.

Suppose on the contrary that Z_2 contains v_1 or $\tilde{v}_1$. Then, by Lemmas 5.1.8–5.1.9, Z begins with a palindromic path from v_1 to $\tilde{v}_1$, and this palindromic path begins with $Y = Z_1 v_{k-1}^{\epsilon_{k-1}}$ (and hence ends with $\tilde{Y}$) by the conditions on Z under the induction hypothesis. More explicitly, we have:

$$Z = \overbrace{\underbrace{v_1 v_2^{\epsilon_2} \cdots v_{k-2} v_{k-1}^{\epsilon_{k-1}}}_{Y} \cdots \underbrace{v_{k-1}^{-\epsilon_{k-1}} \tilde{v}_{k-2} \cdots v_2^{-\epsilon_2} \tilde{v}_1}_{\tilde{Y}}}^{\text{palindromic}} \underbrace{\cdots v_{k-1} v_k^{\epsilon_k}}_{Z_5}.$$

Hence, as v_{k-1} and $\tilde{v}_{k-1}$ are not interior vertices of Y (by the induction hypothesis), the subpath $\tilde{Y}Z_5$ begins with a palindromic path from $v_{k-1}^{-\epsilon_k}$ to $v_{k-1}^{\epsilon_{k-1}}$, and this palindromic path begins with $\tilde{Y}$ (and hence ends with Y), by Lemmas 5.1.8–5.1.9. But then Z ends with a shorter path from Y to $v_{k-1}v_k^{\epsilon_k}$, contradicting the minimality of Z.

We conclude that the subpath $Z_2 = v_{k-1}^{\epsilon_{k-1}} \cdots v_{k-1}$ does not contain v_j, $\tilde{v}_j$ for all j with $1 \leq j \leq k$, $j \neq i$ (i.e., the path Z satisfies condition *ii*)), and the proof is thus complete. $\square$

Lemma 5.1.13. *Suppose* $\mathbf{w}$ *is a recurrent rich infinite word. Then the super reduced Rauzy graph* $\Gamma''_n(\mathbf{w})$ *is a tree for all* $n \in \mathbb{N}_+$.

Proof. First recall that for all n, $\Gamma''_n(\mathbf{w})$ is connected (by the recurrence property of $\mathbf{w}$). Moreover, Lemma 5.1.11 tells us that if two distinct vertices in $\Gamma''_n(\mathbf{w})$ are joined by an edge, then this edge is unique (and corresponds to a simple path and its reversal). It remains to show that $\Gamma''_n(\mathbf{w})$ does not contain any *cycle* (i.e., does not contain a chain linking a vertex with itself).

Suppose on the contrary that $\Gamma''_n(\mathbf{w})$ contains a cycle for some n. Then $\Gamma''_n(\mathbf{w})$ must contain at least three distinct vertices: $[v_1]$, $[v_2]$, $\ldots$, $[v_s]$, $s \geq 3$, and a cycle of the following form:

$$[v_1] \text{—} [v_2] \text{—} \cdots \text{—} [v_k] \text{—} [v_1] \quad \text{for some } k \text{ with } 3 \leq k \leq s. \tag{5.6}$$

We thus deduce from Lemma 5.1.12 that the reduced Rauzy graph $\Gamma'_n(\mathbf{w})$ contains a path from v_1 to $v_1^{\epsilon_1}$ of the form:

$$P = v_1 v_2^{\epsilon_2} \cdots v_2 v_3^{\epsilon_3} \cdots v_{k-2} v_{k-1}^{\epsilon_{k-1}} \cdots v_{k-1} v_k^{\epsilon_k} \cdots v_k v_1^{\epsilon_1},$$

where for all $i = 2, \ldots, k$, the subpath $v_i^{\epsilon_i} \cdots v_i$ (which may consist of only the single vertex $v_i^{\epsilon_i}$) does not contain v_j, $\tilde{v}_j$ for all j with $1 \leq j \leq k$, $j \neq i$. (Note that P corresponds to the cycle given in (5.6).)

First suppose that v_1 is a palindrome. In this case, as neither v_1 nor $\tilde{v}_1$ is an interior vertex of P, it must be a palindromic path by Lemma 5.1.8. But then $v_k = v_2^{-\epsilon_2}$, a contradiction (as $k \geq 3$).

Now suppose that v_1 is not a palindrome. If $v_1^{\epsilon_1} = \tilde{v}_1$, then we deduce (as above, using Lemma 5.1.8) that the path P must be palindromic, yielding a contradiction. On the other hand, if $v_1^{\epsilon} = v_1$, then, by Lemma 5.1.9, the path P must pass through $\tilde{v}_1$, a contradiction.

Thus $\Gamma''_n(\mathbf{w})$ is a tree. $\square$

This concludes our proof of the "(I) $\Rightarrow$ (II)" part of Theorem 5.1.1.

(II) implies (I)

Conversely, suppose $\mathbf{w}$ is an infinite word with $F(\mathbf{w})$ closed under reversal and satisfying equality (II). Then $\mathbf{w}$ satisfies conditions 1) and 2) of Proposition 5.1.10.

Now, arguing by contradiction, suppose $\mathbf{w}$ does not satisfy property (I) (i.e., $\mathbf{w}$ is not rich). Then there exists a palindromic factor p that has a non-palindromic complete return u in $\mathbf{w}$; in particular, we have $u = pqavb\tilde{q}p$ for some words q, v (possibly empty) and letters a, b, with $a \neq b$. So the words pqa, $b\tilde{q}p$ and their reversals $a\tilde{q}p$, pqb are factors of $\mathbf{w}$. Thus pq (resp. $\tilde{q}p$) is a right-special (resp. left-special) factor of $\mathbf{w}$. Hence, if u does not contain any other special factors, then u forms the label of a non-palindromic simple path beginning with pq and ending with $\tilde{q}p$. But this contradicts condition 1) of Proposition 5.1.10. Therefore u must contain other special factors of length $n := |pq|$, besides pq and $\tilde{q}p$. In particular, u begins with the label of a simple path of order n beginning with pq and ending with another special factor s_1 of length n. Similarly, u ends with the label of a simple path of order n beginning with a special factor s_2 of length n and ending with $\tilde{q}p$. Moreover, since u is a complete return to p, neither s_1 nor s_2 is equal to pq or $\tilde{q}p$ (otherwise p occurs as an interior factor of u). Thus, in the super reduced Rauzy graph $\Gamma''_n(\mathbf{w})$, there is an edge between the vertex $[pq]$ and each of the vertices $[s_1]$ and $[s_2]$. In particular, there exists a path of the form: $[s_1]$—$[pq]$—$[s_2]$. Furthermore, as u contains a factor that begins with s_1 and ends with s_2 and contains no occurrence of pq or $\tilde{q}p$, there also exists a chain (or possibly just an edge) linking $[s_1]$ and $[s_2]$ that does not contain the vertex $[pq]$. Thus, if $\{s_1, \tilde{s}_1\} \neq \{s_2, \tilde{s}_2\}$, then we see that $\Gamma''_n(\mathbf{w})$ contains a cycle, contradicting condition 2) of Proposition 5.1.10. On the other hand, if $\{s_1, \tilde{s}_1\} = \{s_2, \tilde{s}_2\}$, then there are at least two edges joining the vertices $[s_1]$ and $[pq]$. Indeed, there exists a simple path P_1 from pq to s_1 and there also exists a simple path P_2 either from s_1 to $\tilde{q}p$ or from $\tilde{s}_1$ to $\tilde{q}p$. By closure under reversal, the reversals $\tilde{P}_1$, $\tilde{P}_2$ of the respective simple paths P_1, P_2 also exist. Moreover, none of these four simple paths coincide. Certainly, $P_1 \neq P_2$, $P_1 \neq \tilde{P}_1$, and $P_2 \neq \tilde{P}_2$ as neither s_1 nor $\tilde{s}_1$ is equal to pq or $\tilde{q}p$, and $P_1 \neq \tilde{P}_2$ as the second vertex in P_1 ends with the letter a, whereas the second vertex in the path $\tilde{P}_2$ ends with the letter $b \neq a$. So $\Gamma''_n(\mathbf{w})$ is not a tree, contradicting condition 2) of Proposition 5.1.10. This concludes our proof of Theorem 5.1.1. □

5.2 Factors of rich words

The following proposition collects together all of the characteristic properties of rich words that were previously established in [29] and in Proposition 5.1.2.

Proposition 5.2.1. *For any finite or infinite word w, the following conditions are equivalent:*

i) w is rich;

ii) every prefix of w has a unioccurrent palindromic suffix (and equivalently, when w is finite, every suffix of w has a unioccurrent palindromic prefix);

iii) every factor u of w contains exactly $|u| + 1$ distinct palindromes;

iv) for each factor u of w, every prefix (resp. suffix) of u has a unioccurrent palindromic suffix (resp. prefix);

v) for each palindromic factor p of w, every complete return to p in w is a palindrome.

Remark. The equivalences: i) $\Leftrightarrow$ ii), i) $\Leftrightarrow$ iii), and i) $\Leftrightarrow$ iv) were proved in [29].

Explicit characterizations of periodic rich infinite words and recurrent *balanced* rich infinite words have also been established in [38]. In the preceding section, the following connection between palindromic richness and complexity has been proved,

Proposition 5.2.2. *For any infinite word* $\mathbf{w}$ *whose set of factors is closed under reversal, the following conditions are equivalent:*

- *all complete returns to palindromes are palindromes;*
- $\mathcal{P}(n) + \mathcal{P}(n+1) = \mathcal{C}(n+1) - \mathcal{C}(n) + 2$ *for all* $n \in \mathbb{N}$,

where $\mathcal{P}$ *(resp.* $\mathcal{C}$*) denotes the* palindromic complexity *(resp.* factor complexity*) function of* $\mathbf{w}$*, which counts the number of distinct palindromic factors (resp. factors) of each length in* $\mathbf{w}$*.*

From the perspective of richness, the above result can be viewed as a characterization of *recurrent* rich infinite words since any rich infinite word is recurrent if and only if its set of factors is closed under reversal (see [38]). Interestingly, the proof of Proposition 5.2.2 relied upon another characterization of rich words, stated below.

Proposition 5.2.3. *A finite or infinite word* w *is rich if and only if, for each factor* v *of* w*, every factor of* w *beginning with* v *and ending with* $\tilde{v}$ *and containing no other occurrences of* v *nor of* $\tilde{v}$ *is a palindrome.*

In this section, we establish yet another interesting characteristic property of rich words. Our main results are the following two theorems.

Theorem 5.2.4. *For any finite or infinite word* w*, the following conditions are equivalent:*

(A) w *is rich;*

(B) each non-palindromic factor u *of* w *is uniquely determined by a pair* (p, q) *of palindromes such that* p *and* q *are not factors of each other and* p *(resp.* q*) is the longest palindromic prefix (resp. suffix) of* u*.*

Theorem 5.2.5. *A finite or infinite word* w *is rich if and only if each factor of* w *is uniquely determined by its longest palindromic prefix and its longest palindromic suffix.*

By contrast, a rich word is *not* uniquely determined by its longest palindromic prefix and suffix. For example, consider the words acb and adb where a, b, c, d are mutually distinct letters. These two words are rich with the same longest palindromic prefix (namely a) and the same longest palindromic suffix (namely b).

5.2.1 Proofs of Propositions 5.2.1 and 5.2.3

For the sake of completeness, we will first provide simple proofs of the characteristic properties stated in Propositions 5.2.1 and 5.2.3.

Proof of Proposition 5.2.1. We begin by proving the equivalence of properties i) and ii).

i) $\Leftrightarrow$ ii): Let $P(w)$ denote the number of distinct palindromic factors of w. For any word u and letter x, we have

$$P(ux) = \begin{cases} P(u) & \text{if } ux \text{ does not have a unioccurrent palindromic suffix,} \\ P(u)+1 & \text{if } ux \text{ has a unioccurrent palindromic suffix.} \end{cases}$$

Therefore, by induction (with $P(\varepsilon) = 1$), it follows that $P(w)$ is precisely the number of prefixes of w that have a unioccurrent palindromic suffix. In particular $P(w) \leq |w| + 1$, and moreover we see that $P(w) = |w| + 1$ (i.e., w is rich) if and only if each prefix of w has a unioccurrent palindromic suffix. Similarly, when w is finite, we deduce that w is rich if and only if each suffix of w has a unioccurrent palindromic prefix.

ii) $\Rightarrow$ iii): Suppose w satisfies property ii) (i.e., w is rich) and let u be any factor of w. Then $w = vuv'$ for some words v, v' where v is finite, and v' is finite or infinite depending on w. By property ii), every prefix of vu has a unioccurrent palindromic suffix, and so again by ii), every suffix of u have a unioccurrent palindromic prefix. Thus, by the equivalence of properties i) and ii), u is rich, i.e., u has exactly $|u| + 1$ distinct palindromic factors.

iii) $\Rightarrow$ iv): Suppose w satisfies property iii). Then every factor of w is rich. Hence, for each factor u of w, every prefix (resp. suffix) of u has a unioccurrent palindromic suffix (resp. prefix), by the equivalence of properties i) and ii).

iv) $\Rightarrow$ v): Suppose to the contrary that property v) does not hold for w satisfying property iv). Then w contains a non-palindromic complete return r to a palindrome p. We deduce that $r = pup$ for some non-palindromic word u. Indeed, since r is not a palindrome, $r \neq pp$ and the two occurrences of p in r cannot overlap; otherwise there exists a non-empty word v such that $r = pv^{-1}p$, in which case $p = vf = gv = \tilde{v}\tilde{g} = \tilde{p}$ for some words f, g. Whence $v = \tilde{v}$ and $r = g\tilde{v}\tilde{g} = gv\tilde{g}$, a palindrome. Now, we easily see that p is the longest palindromic suffix of r; otherwise p would occur in the interior of r as a prefix of a longer palindromic suffix of r. But then r does not have a unioccurrent palindromic suffix (as p is also a prefix of r), a contradiction.

v) $\Rightarrow$ i): Suppose not. Let u be a factor of w of minimal length satisfying property v) and not rich. Since all words of length 3 or less are rich (easy to check), we may write $u = xvy$ with x, y letters and v a word of length at least 2. By the minimality of u, xv is rich and by the equivalence of i) and ii), the longest palindromic suffix p of u occurs more than once in u. Hence, by property v), we reach a contradiction to the maximality of p. $\square$

Proof of Proposition 5.2.3. ONLY IF: Consider any factor v of w and let u be a factor of w beginning with v and ending with $\tilde{v}$ and containing neither v nor $\tilde{v}$ as an interior factor. If v is a palindrome, then either $u = v = \tilde{v}$ (in which case u is clearly a palindrome), or u is a complete return to v in w, and hence u is (again) a palindrome by Proposition 5.2.1.

Now assume that v is not a palindrome. Suppose by way of contradiction that u is not a palindrome and let p be the longest palindromic suffix of u (which is unioccurrent in u by richness). Then $|p| < |u|$ as u is not a palindrome. If $|p| > |v|$, then $\tilde{v}$ is a proper suffix of p, and hence v is a proper prefix of p. But then v is an interior factor of u, a contradiction. On the other hand, if $|p| \leq |v|$, then $|p| \neq |v|$ and p is a proper suffix of $\tilde{v}$ (as $\tilde{v}$ is not a palindrome), and hence p is a proper prefix of v. Thus p is both a prefix and a suffix of u; in particular p is not unioccurrent in u, a contradiction.

IF: The given conditions tell us that any complete return to a palindromic factor v $(= \tilde{v})$ of w is a palindrome. Hence w is rich by Proposition 5.2.1. $\square$

5.2.2 Proof of Theorem 5.2.4

We will now prove our first main theorem. The following two lemmas establish that (A) implies (B).

Lemma 5.2.6. *Suppose w is a finite or infinite rich word and let u be any non-palindromic factor of w with longest palindromic prefix p and longest palindromic suffix q. Then $p \neq q$, and p and q are not proper factors of each other.*

Proof. By Proposition 5.2.1, p and q are unioccurrent factors of u. Thus, since u is not a palindrome (and hence $|u| > \max\{|p|, |q|\}$), it follows immediately that $p \neq q$, and p and q are not proper factors of each other. $\square$

Lemma 5.2.7. *Suppose w is a finite or infinite rich word. If u and v are factors of w with the same longest palindromic prefix p and the same longest palindromic suffix q, then $u = v$.*

Proof. We first observe that if u or v is a palindrome, then $u = p = q = v$. So let us now assume that neither u nor v is a palindrome.

Suppose to the contrary that $u \neq v$. Then u and v are clearly not factors of each other since neither u nor v is equal to p or q, and p and q are unioccurrent in each of u and v (by Proposition 5.2.1). Let z be a factor of w of minimal length containing both u and v. As u and v are not factors of each other, we may assume without loss of generality that z begins with u and ends with v. Then z contains *at least* two distinct occurrences of p (as a prefix of each of u and v). In particular, z begins with a complete return r_1 to p with $|r_1| > |u|$ because p is unioccurrent in u by Proposition 5.2.1. Moreover, r_1 is a palindrome by the richness of w, and hence r_1 ends with $\tilde{u}$ since u is a proper prefix of r_1. Similarly, z ends with a complete return r_2 to q with $|r_2| > |v|$ since q is unioccurrent in v by Proposition 5.2.1. Hence, since r_2 is a palindrome (by the richness of w) and v is a proper suffix of r_2, it follows that r_2 begins with $\tilde{v}$. So we have shown that $\tilde{u}$ and $\tilde{v}$ are (distinct) interior factors of z.

Let us first suppose that an occurrence of $\tilde{v}$ is followed by an occurrence of $\tilde{u}$ in z (i.e., z has an interior factor beginning with $\tilde{v}$ and ending with $\tilde{u}$). Then, since q is a unioccurrent prefix of each of the (distinct) factors $\tilde{v}$ and $\tilde{u}$, we deduce that z contains (as an interior factor) a complete return r_3 to q beginning with $\tilde{v}$. In particular, as r_3 is a palindrome (by richness), r_3 ends with v. Thus, z has a proper prefix beginning with u and ending with v, contradicting the minimality of z. On the other hand, if z has an interior factor beginning with $\tilde{u}$ and ending with $\tilde{v}$, then using the same reasoning as above, we deduce that z has a proper suffix beginning with u and ending with v. But again, this contradicts the minimality of z; whence $u = v$. $\square$

The proof of "(A) $\Rightarrow$ (B)" is now complete. The next lemma proves that

(B) implies (A).

Lemma 5.2.8. *Suppose w is a finite or infinite word with the property that each non-palindromic factor u of w is uniquely determined by a pair (p, q) of distinct palindromes such that p and q are not factors of each other and p (resp. q) is the longest palindromic prefix (resp. suffix) of u. Then w is rich.*

Proof. To prove that w is rich, it suffices to show that each prefix of w has a unioccurrent palindromic suffix (see Proposition 5.2.1).

Let u be any prefix of w and let q be the longest palindromic suffix of u. We first observe that if u is a palindrome then $u = q$, and hence q is unioccurrent in u. Now let us suppose that u is not a palindrome and let p be the longest palindromic prefix of u. If q is not unioccurrent in u, then, as p and q are not factors of each other (by the given property of w), we deduce that u has a *proper* factor v beginning with p and ending with q and containing neither p nor q as an interior factor. Moreover, we observe that p is the longest palindromic prefix of v; otherwise p would occur in the interior of v (as a suffix of a longer palindromic prefix of v). Similarly, we deduce that q is the longest palindromic suffix of v. So v has the same longest palindromic prefix and the same longest palindromic suffix as u, a contradiction. Whence q is unioccurrent in u. This completes the proof of the lemma. □

5.2.3 Proof of Theorem 5.2.5

Lemma 5.2.7 proves that each factor of a rich word is uniquely determined by its longest palindromic prefix and its longest palindromic suffix.

Conversely, suppose w is a finite or infinite word with the property that each factor of w is uniquely determined by its longest palindromic prefix and its longest palindromic suffix. To prove that w is rich, we could use very similar reasoning as in the proof of Lemma 5.2.8. But for the sake of interest, we give a slightly different proof. Specifically, we show that all complete returns to any palindromic factor of w are palindromes; whence w is rich by Proposition 5.2.1.

Let us suppose to the contrary that w contains a non-palindromic complete return r to a palindromic factor p. Then $r = pvp$ for some non-palindromic word v (as already observed in the proof vi) $\Rightarrow$ v) in Proposition 5.2.1). We easily see that p is both the longest palindromic prefix and the longest palindromic suffix of r; otherwise p would occur in the interior of r as a suffix of a longer palindromic prefix of r, or as a prefix of a longer palindromic suffix of r. As $r \neq p$, we have reached a contradiction to the fact that p is the *only* factor of w having itself as both its longest palindromic prefix and its longest palindromic suffix. Thus, all complete returns to p in w are palindromes. This completes the proof of Theorem 5.2.5. □

5.3 A few consequences and remarks

From Theorem 5.1.1, we easily deduce that property (I) is equivalent to equality (II) for any uniformly recurrent infinite word. Indeed, equality (II) implies the existence of arbitrarily long palindromes since $\mathcal{P}(n) + \mathcal{P}(n+1) \geq 2$ for all n, so together with uniform recurrence one can readily show that factors are closed under reversal; hence property (I) holds by Theorem 5.1.1. Conversely, richness (property (I)) together with uniform recurrence implies closure under reversal by Remark 5.1.1, and hence equality (II) holds.

Question: *In the statement of Theorem 5.1.1, can the hypothesis of factors being closed under reversal be replaced by the weaker hypothesis of recurrence?*

As above, it follows directly from Theorem 5.1.1 and Remark 5.1.1 that for any recurrent infinite word $\mathbf{w}$, if $\mathbf{w}$ satisfies property (I) (i.e., if $\mathbf{w}$ is rich, and hence has factors closed under reversal), then equality (II) holds. However, to prove the converse using our methods, one would need to know that any recurrent infinite word satisfying equality (II) has factors closed under reversal. We could not find a proof of this claim nor could we find a counter-example. Let us point out that whilst uniform recurrence and the existence of arbitrarily long palindromes imply closure under reversal, this is not true in the case of recurrence only. For instance, consider the following infinite word:

$$\mathbf{s} = bca^2bca^3bca^2bca^4bca^2bca^3bca^2bca^5bc\cdots,$$

which is the limit as n goes to infinity of the sequence $(s_n)_{n\geq 1}$ of finite words defined by:

$$s_1 = bc \quad \text{and} \quad s_n = s_{n-1}a^n s_{n-1} \quad \text{for } n > 1.$$

This infinite word is clearly recurrent (but not uniformly recurrent) and contains arbitrarily long palindromes, but its set of factors is not closed under reversal. (Note that s is not rich and does not satisfy equality (II).) If one could show that recurrence together with equality (II) implies arbitrarily long palindromic prefixes, this would be enough to prove that factors are closed under reversal.

In the context of finite words w, the hypothesis of factors being closed under reversal can be replaced by the requirement that w is a palindrome. Indeed, all we really need is the super reduced Rauzy graph to be connected, which is true for palindromes.

Theorem 5.3.1. *For any palindrome w, the following properties are equivalent:*

i) *w contains $|w|+1$ distinct palindromes;*

ii) *all complete returns to palindromes in w are palindromes;*

iii) $\mathcal{P}(i)+\mathcal{P}(i+1)=\mathcal{C}(i+1)-\mathcal{C}(i)+2$ *for all i with $0 \leq i \leq |w|$.* □

We now prove two easy consequences of Theorem 5.1.1.

Corollary 5.3.2. *Suppose* $\mathbf{w}$ *is a recurrent rich infinite word. Then the following properties hold.*

i) $\mathbf{w}$ *is (purely) periodic if and only if $\mathcal{P}(n)+\mathcal{P}(n+1)=2$ for some n.*

ii) $(\mathcal{P}(n))_{n\geq 1}$ *is eventually periodic with period 2 if and only if there exist non-negative integers K, L, N such that $\mathcal{C}(n)=Kn+L$ for all $n \geq N$.*

Proof. Suppose $\mathbf{w}$ is a recurrent rich infinite word. Then $\mathcal{P}(n)+\mathcal{P}(n+1)=\mathcal{C}(n+1)-\mathcal{C}(n)+2$ for all n, by Theorem 5.1.1 and Remark 5.1.1.

i): If $\mathcal{P}(n)+\mathcal{P}(n+1)=2$ for some n, then $\mathcal{C}(n+1)=\mathcal{C}(n)$, and hence $\mathbf{w}$ is eventually periodic; in particular, $\mathbf{w}$ must be (purely) periodic as it is recurrent. Conversely, if $\mathbf{w}$ is periodic, then $\mathcal{C}(n+1)=\mathcal{C}(n)$ for some n, and hence $\mathcal{P}(n)+\mathcal{P}(n+1)=2$.

ii): The condition on $\mathcal{C}(n)$ implies that for all $n \geq N$, $\mathcal{C}(n+1)-\mathcal{C}(n)=K$, and hence $\mathcal{P}(n)+\mathcal{P}(n+1)=K+2=\mathcal{P}(n+1)+\mathcal{P}(n+2)$. Thus $\mathcal{P}(n)=\mathcal{P}(n+2)$ for all $n \geq N$. Conversely, suppose $(\mathcal{P}(n))_{n\geq 1}$ is eventually periodic with period 2. Then there exists a non-negative integer N such that $\mathcal{P}(n) = \mathcal{P}(n+2)$ for all $n \geq N$. Hence, for all $n \geq N$, $\mathcal{P}(n) + \mathcal{P}(n+1) = \mathcal{C}(n+1) - \mathcal{C}(n) + 2 = \mathcal{P}(n+1) + \mathcal{P}(n+2) = M \geq 2$. Therefore $\mathcal{C}(n+1) - \mathcal{C}(n) = M - 2$ for all $n \geq N$. $\square$

Remark. Item *ii*) of the above corollary can be compared with a result of J. Cassaigne [19], who proved that if $\mathcal{C}(n)$ has linear growth, then $\mathcal{C}(n+1) - \mathcal{C}(n)$ is bounded.

Remark. In [5], Balaži *et al.* remarked: "According to our knowledge, all known examples of infinite words which satisfy the equality $\mathcal{P}(n) + \mathcal{P}(n+1) = \mathcal{C}(n+1) - \mathcal{C}(n) + 2$ for all $n \in \mathbb{N}$ have sublinear factor complexity." Actually, there do exist recurrent rich infinite words with non-sublinear complexity. For instance, the following example from [38]: $abab^2abab^3abab^2abab^4abab^2abab^3abab^2abab^5\cdots$ (which is the fixed point of the morphism: $a \mapsto abab$, $b \mapsto b$) is a recurrent rich infinite word and its complexity $\mathcal{C}(n)$ grows quadratically with n. Another example that was indicated to us by J. Cassaigne is the fixed point of $a \mapsto aab$, $b \mapsto b$:

$$aabaabbaabaabbbaabaabbaabaabbbbaabaabbaabaabbbaabaabbaabaabbbbb\cdots.$$

It is a recurrent rich infinite word and its complexity is equivalent to $n^2/2$. More precisely, $\mathcal{P}(n) + \mathcal{P}(n+1) - 2 = \mathcal{C}(n+1) - \mathcal{C}(n) = n + 1 - \sharp\{k > 0 \mid 2^k + k - 2 < n\}$.

In [29], X. Droubay *et al.* showed that the family of *Episturmian words* (e.g., see [29, 43, 37]), which includes the well-known *Sturmian words*, comprises a special class of uniformly recurrent rich infinite words. Specifically, they proved that if an infinite word $\mathbf{w}$ is Episturmian, then any factor u of $\mathbf{w}$ contains exactly $|u| + 1$ distinct palindromic factors (see [29, Cor. 2]). An alternative proof of the richness of Episturmian words can be found in the paper [3] where the fourth author, together with V. Anne and I. Zorca, proved that for Episturmian words, all complete returns to palindromes are palindromes. (A shorter proof of this fact is also given in [13].) More recently, P. Baláži *et al.* [5] showed that all *strict* Episturmian words (i.e., *Arnoux-Rauzy sequences* [4, 53]) satisfy $\mathcal{P}(n) + \mathcal{P}(n+1) = \mathcal{C}(n+1) - \mathcal{C}(n) + 2$ for all n. This fact, together with Theorem 5.1.1, provides yet another proof that *all* Episturmian words are rich (since any factor of an Episturmian word is a factor of some strict Episturmian word).

Sturmian words are exactly the aperiodic Episturmian words over a 2-letter alphabet. They have complexity $n+1$ for each n and are characterized by their palindromic complexity: any Sturmian word has $\mathcal{P}(n) = 1$ whenever n is even and $\mathcal{P}(n) = 2$ whenever n is odd (see [30]). From these observations, one can readily check that Sturmian words satisfy equality (II) (and hence they are rich).

We can now say even more: the set of factors of all Sturmian words satisfies equality (II). To show this, we first recall that F. Mignosi [48] proved that, for any $n \geq 0$, the number $c(n)$ of *finite Sturmian words* of length n is given by

$$c(n) = 1 + \sum_{i=1}^{n}(n+1-i)\phi(i),$$

where ϕ is *Euler's totient function*. More recently, in [26], the second author together with A. de Luca proved that for any $n \geq 0$, the number $p(n)$ of Sturmian palindromes of length n is given by

$$p(n) = 1 + \sum_{i=0}^{\lceil n/2\rceil - 1} \phi(n-2i).$$

Equivalently, for any $n \geq 0$,

$$p(2n) = 1 + \sum_{i=1}^{n}\phi(2i) \quad \text{and} \quad p(2n+1) = 1 + \sum_{i=0}^{n}\phi(2i+1).$$

Thus, for all $n \geq 0$,

$$p(2n) + p(2n+1) = 2 + \sum_{i=1}^{n}\{\phi(2i) + \phi(2i+1)\} + 2 = \sum_{i=1}^{2n+1}\phi(i) + 2,$$

and

$$\begin{aligned} c(2n+1) - c(2n) + 2 &= \sum_{i=1}^{2n+1}(2n+2-i)\phi(i) - \sum_{i=1}^{2n}(2n+1-i)\phi(i) + 2 \\ &= \phi(2n+1) + \sum_{i=1}^{2n}\phi(i) + 2 \\ &= \sum_{i=1}^{2n+1}\phi(i) + 2 = p(2n) + p(2n+1). \end{aligned}$$

Bibliography

[1] J.-P. Allouche, M. Baake, J. Cassaigne, and D. Damanik. Palindrome complexity. *Theoretical Computer Science*, 292:9–31, 2003.

[2] P. Ambrož, C. Frougny, Z. Masáková, and E. Pelantová. Palindromic complexity of infinite words associated with simple Parry numbers. *Ann. Inst. Fourier (Grenoble)*, 56:2131–2160, 2006.

[3] V. Anne, L. Q. Zamboni, and I. Zorca. Palindromes and pseudo-palindromes in episturmian and pseudo-palindromic infinite words. In S. Brlek and C. Reutenauer, editors, *Words 2005*, number 36 in Publications du LaCIM, pages 91–100, 2005.

[4] P. Arnoux and G. Rauzy. Représentation géométrique de suites de complexité $2n+1$. *Bulletin de la Société Mathématique de France*, 119:199–215, 1991.

[5] P. Baláži, Z. Masáková, and E. Pelantová. Factor versus palindromic complexity of uniformly recurrent infinite words. *Theoretical Computer Science*, 380:266–275, 2007.

[6] J. Berstel and A. de Luca. Sturmian words, Lyndon words and trees. *Theoretical Computer Science*, 178:171–203, 1997.

[7] J. Berstel and D. Perrin. *Theory of Codes*. Academic Press, New York, 1985.

[8] J. Berstel and P. Séébold. Sturmian words. In M. Lothaire, editor, *Algebraic Combinatorics on Words*. Cambridge University Press, Cambridge UK, 2002. Chapter 2.

[9] J.-P. Borel and C. Reutenauer. Palindromic factors of billiard words. *Theoretical Computer Science*, 340:334–348, 2005.

[10] S. Brlek, S. Hamel, M. Nivat, and C. Reutenauer. On the palindromic complexity of infinite words. *International Journal of Foundations of Computer Science*, 15:293–306, 2004.

[11] M. Bucci, A. de Luca, and A. De Luca. Characteristic morphisms of generalized episturmian words. *Theoretical Computer Science*, to appear, 2008.

[12] M. Bucci, A. de Luca, A. De Luca, and L. Q. Zamboni. On different generalizations of episturmian words. *Theoretical Computer Science*, 393:23–36, 2008.

[13] M. Bucci, A. de Luca, A. De Luca, and L. Q. Zamboni. On some problems related to palindrome closure. *Theoretical Informatics and Applications*, DOI: 10.1051/ita:2007064, 2008. to appear.

[14] M. Bucci, A. de Luca, A. De Luca, and L. Q. Zamboni. On θ-episturmian words. *European Journal of Combinatorics*, to appear, 2008.

[15] M. Bucci, A. De Luca, A. Glen, and L.Q. Zamboni. A new characteristic property of rich words. *Theoretical Computer Science*, To appear, 2008.

[16] M. Bucci, A. De Luca, A. Glen, and L.Q. Zamboni. A connection between palindromic and factor complexity using return words. *Adv. in Appl. Math.*, 42:60–74, 2009.

[17] A. Carpi. Private communication, 2005.

[18] A. Carpi and A. de Luca. Codes of central Sturmian words. *Theoretical Computer Science*, 340:220–239, 2005.

[19] J. Cassaigne. Special factors of sequences with linear subword complexity. In *Developments in Language Theory II*, pages 25–34. World Scientific, Singapore, 1996.

[20] W.-F. Chuan. Moments of conjugacy classes of binary words. *Theoretical Computer Science*, 310:273–285, 2004.

[21] P. M. Cohn. *Algebra*, volume 1–2. John Wiley & Sons, Chichester, 2nd edition, 1982.

[22] F. D'Alessandro. A combinatorial problem on trapezoidal words. *Theoretical Computer Science*, 273:11–33, 2002.

[23] A. de Luca. On some combinatorial problems in free monoids. *Discrete Mathematics*, 38:207–225, 1982.

[24] A. de Luca. Sturmian words: structure, combinatorics, and their arithmetics. *Theoretical Computer Science*, 183:45–82, 1997.

[25] A. de Luca. On the combinatorics of finite words. *Theoretical Computer Science*, 218:13–39, 1999.

[26] A. de Luca and A De Luca. Combinatorial properties of Sturmian palindromes. *Internat. J. Found. Comput. Sci.*, 17:557–573, 2006.

[27] A. de Luca and A. De Luca. Pseudopalindrome closure operators in free monoids. *Theoretical Computer Science*, 362:282–300, 2006.

[28] A. de Luca and F. Mignosi. Some combinatorial properties of Sturmian words. *Theoretical Computer Science*, 136:361–385, 1994.

[29] X. Droubay, J. Justin, and G. Pirillo. Episturmian words and some constructions of de Luca and Rauzy. *Theoretical Computer Science*, 255:539–553, 2001.

[30] X. Droubay and G. Pirillo. Palindromes and Sturmian words. *Theoretical Computer Science*, 223:73–85, 1999.

[31] F. Durand. A characterization of substitutive sequences using return words. *Discrete Mathematics*, 179:89–101, 1998.

[32] F. Durand. A generalization of cobham's theorem. *Theory Comput. Syst.*, 31:169–185, 1998.

[33] F. Durand. Linearly recurrent subshifts have a finite number of non-periodic subshift factors. *Ergodic Theory Dynam. Systems*, 19:953–993, 1999.

[34] S. Ferenczi, C. Mauduit, and A. Nogueira. Substitutional dynamical systems: algebraic characterization of eigenvalues. *Ann. Sci. École Norm. Sup.*, 29:519–533, 1995.

[35] S. Fischler. Palindromic prefixes and episturmian words. *J. Combin. Theory Ser. A*, 113:1281–1304, 2006.

[36] S. Fischler. Palindromic prefixes and diophantine approximation. *Monatsh. Math.*, 151:11–37, 2007.

[37] A. Glen and J. Justin. Episturmian words: a survey. *Preprint*, arXiv:0801.1655, 2007.

[38] A. Glen, J. Justin, S. Widmer, and L.Q. Zamboni. Palindromic richness. *European Journal of Combinatorics*, doi:10.1016/j.ejc.2008.04.006, 2008.

[39] T. Harju and D. Nowotka. Minimal Duval extensions. *Internat. J. Found. Comput. Sci.*, 15:349–354, 2004.

[40] C. Holton and L.Q. Zamboni. Descendants of primitive substitutions. *Theory Comput. Syst.*, 32:133–157, 1999.

[41] O. Jenkinson and L. Q. Zamboni. Characterisations of balanced words via orderings. *Theoretical Computer Science*, 310:247–271, 2004.

[42] J. Justin. Episturmian morphisms and a Galois theorem on continued fractions. *Theoretical Informatics and Applications*, 39:207–215, 2005.

[43] J. Justin and G. Pirillo. Episturmian words and episturmian morphisms. *Theoretical Computer Science*, 276:281–313, 2002.

[44] M. Lothaire. *Combinatorics on Words*. Addison-Wesley, Reading MA, 1983.

[45] M. Lothaire. *Algebraic Combinatorics on Words*. Cambridge University Press, 2002.

[46] M. Lothaire. *Applied Combinatorics on Words*. Cambridge University Press, Cambridge UK, 2005.

[47] S. Mac Lane. *Categories for the Working Mathematician.* Springer-Verlag, New York, 2nd edition, 1998.

[48] F. Mignosi. On the number of factors of Sturmian words. *Theoretical Computer Science*, 82:71–84, 1991.

[49] F. Mignosi and L. Q. Zamboni. A note on a conjecture of Duval and Sturmian words. *Theoret. Inform. Appl.*, 36:1–3, 2002.

[50] M. Morse and G. A. Hedlund. Symbolic dynamics. *American Journal of Mathematics*, 60:815–866, 1938.

[51] M. Morse and G. A. Hedlund. Symbolic dynamics II. Sturmian trajectories. *American Journal of Mathematics*, 62:1–42, 1940.

[52] G. Pirillo. A curious characteristic property of standard Sturmian words. In G.-C. Rota, H. Crapo, and D. Senato, editors, *Algebraic combinatorics and computer science*, pages 541–546. Springer Italia, Milan, 2001.

[53] G. Rauzy. Suites à termes dans un alphabet fini. *Sémin. Théorie des Nombres*, 25:16, 1982–1983.

[54] A. Siegel. Pure discrete spectrum dynamical systems and periodic tiling associated with a substitution. *Ann. Inst. Fourier (Grenoble)*, 54:341–381, 2004.

[55] N. J. A. Sloane. *The On-Line Encyclopedia of Integer Sequences.* AT&T Research, `http://www.research.att.com/~njas/sequences/`.